

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Deep learning, perceptual modeling, and human-in-the-loop stimulus optimization for artificial vision

Permalink

<https://escholarship.org/uc/item/70q0k9kg>

ISBN

9798297686069

Author

Granley, Jacob

Publication Date

2025-09-10

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

**Deep learning, perceptual modeling, and
human-in-the-loop stimulus optimization for artificial
vision**

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Computer Science

by

Jacob Granley

Committee in charge:

Professor Michael Beyeler, Chair
Professor Miguel Eckstein
Professor Ambuj Singh

September 2025

The Dissertation of Jacob Granley is approved.

Professor Miguel Eckstein

Professor Ambuj Singh

Professor Michael Beyeler, Committee Chair

September 2025

Deep learning, perceptual modeling, and human-in-the-loop stimulus optimization for
artificial vision

Copyright © 2025

by

Jacob Granley

To my Mom. Your love, support, and relentless dedication to enabling and enriching my education are the only reason any of this was possible.

Acknowledgements

First and foremost, I need to express my deepest gratitude to my advisor, Michael Beyeler. His guidance and mentorship have been absolutely instrumental in every facet of my PhD research, and the friendly and productive research environment he's fostered has made me incredibly grateful to have been a part of the Bionic Vision Lab.

I would also like to thank my committee members, Miguel Eckstein and Ambuj Singh, for their time and feedback throughout the many iterations of these projects. Additionally, I am thankful to Eduardo Fernandez for being a gracious and welcoming host during my months of research in his lab.

I am incredibly lucky to have had such wonderful lab mates. Their passion, intellectual discussions, and friendship in and outside the lab have not only enhanced my research, but have truly made Santa Barbara feel like home.

To the many other friends, family members, colleagues, past mentors, and teachers that have made this possible, thank you all for your support, encouragement, friendship, and inspiration.

Last, but certainly not least, I want to thank my family. To my parents, Brad and Kathy, and my sisters Alyssa and Lily: Your love, support, and sacrifices are the foundation upon which I've built my academic journey – I am indebted beyond what words could ever express.

CURRICULUM VITÆ

Jacob Granley

Education

2025	Ph.D., Computer Science (Expected) University of California, Santa Barbara
2020	M.S., Computer Science Colorado School of Mines
2019	B.S., Computer Science Colorado School of Mines

Publications

Eirini Schoinas, Adyah Rastogi, Anissa Carter, **Jacob Granley**, and Michael Beyeler. “Evaluating Deep Human-in-the-Loop Optimization for Retinal Implants Using Sighted Participants”. In: 2025 47th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE (2025)

Jacob Granley^{*}, Galen Pogoncheff^{*}, Alfonso Rodil, Leili Soo, Lily Marie Turkstra, Lucas Gil Nadolskis, Arantxa Alfaro Saez, Cristina Soto Sanchez, Eduardo Fernandez Jover, and Michael Beyeler. “Beyond Sight: Probing Alignment Between Image Models and Blind V1”. en. In: ICLR 2024 Workshop on Representational Alignment (2024). *Oral Presentation*

Yuchen Hou, Devyani Nanduri, **Jacob Granley**, James D. Weiland, and Michael Beyeler. Axonal stimulation affects the linear summation of single-point perception in three Argus II users. en. In Journal of Neural Engineering (2023)

Jacob Granley, Tristan Fauvel, Matthew Chalk, and Michael Beyeler. “Human-in-the-Loop Optimization for Deep Stimulus Encoding in Visual Prostheses”. en. In: Advances

in Neural Information Processing Systems 36 (2023), pp. 79376–79398

Galen Pogoncheff, **Jacob Granley**, and Michael Beyeler. “Explaining V1 Properties with a Biologically Constrained Deep Learning Architecture.” en. In: Advances in Neural Information Processing Systems 36 (2023), pp. 79376–79398

Jacob Granley, Lucas Relic, and Michael Beyeler. “Hybrid Neural Autoencoders for Stimulus Encoding in Visual and Other Sensory Neuroprostheses”. en. In: Advances in Neural Information Processing Systems 35 (2022), pp. 22671–22685

Jacob Granley, Alexander Riedel, and Michael Beyeler. “Adapting Brain-Like Neural Networks for Modeling Cortical Visual Prostheses”. en. In: Oct. 2022

Jacob Granley and Michael Beyeler. “A Computational Model of Phosphene Appearance for Epiretinal Prostheses”. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). ISSN: 2694-0604. Nov. 2021, pp. 4477–4481

Shuyun Tang, Ziming Qi, **Jacob Granley**, Michael Beyeler. “U-net with hierarchical bottleneck attention for landmark detection in fundus images of the degenerated retina”. In MICCAI Ophthalmic Medial Image Analysis Workshop (2021)

Ahmed Alshehri, **Jacob Granley**, and Chuan Yue. “Attacking and Protecting Tunneled Traffic of Smart Home Devices”. In: CODASPY 2020 - Proceedings of the 10th ACM Conference on Data and Application Security and Privacy. 2020. isbn: 978-1-4503-7107-0

Conference Abstracts

Jacob Granley, Lily M Turkstra, Galen Pogoncheff, Fabrizio Grani, Leili Soo, Alfonso Rodil, Cristina Soto, Thomas C Sprague, Eduardo Fernandez, Michael Beyeler. “Computational methods for extracting neural correlates of working memory and mental imagery

from intracortical recordings in human visual cortex”. Vision Society Sciences (2025), *Oral Presentation*.

Jacob Granley*, Pehuen Moure*, Fabrizio Grani*, Leili Soo, Antonio Lozano, Rocio Lopez-Peco, Adrian Villamarin-Ortiz, Cristina Soto-Sanchez, Shih-Chii Liu, Michael Beyeler, Eduardo Fernandez. “Deep Learning–Based Control of Electrically Evoked Activity in Human Visual Cortex”. The Brain and the Chip Conference (2024), *Oral Presentation*

Jacob Granley, Tristan Fauvel, Matthew Chalk, and Michael Beyeler. “Human-in-the-Loop Optimization Stimulus Encoding in Visual Prostheses”. The Eye and the Chip Conference (2024), *Poster*.

Jacob Granley, Yuchen Hou, Tanya Bhatia, Byron Johnson, Michael Beyeler, “Effects of Stimulus Parameters on Phosphene Appearance in Epiretinal Prostheses”. The Eye and the Chip Conference (2022), *Poster*.

Jacob Granley, Yuchen Hou, Tanya Bhatia, Byron Johnson, Michael Beyeler. “Effects of Stimulus Parameters on Phosphene Appearance in Epiretinal Prostheses”. Investigative Ophthalmology & Visual Science (ARVO) (2022), *Poster*.

Publications - Under Review and In Preparation

Mara Downing, Matthew Peng, **Jacob Granley**, Michael Beyeler, Tevfik Bultan. “Fuzzing the Brain: Discovering Safety Violations in ML-Driven Neurotechnology”. Under review at Advances in Neural Information Processing Systems (2025)

Jacob Granley*, Pehuen Moure*, Fabrizio Grani*, Leili Soo, Antonio Lozano, Rocio Lopez-Peco, Adrian Villamarin-Ortiz, Cristina Soto-Sanchez, Shih-Chii Liu, Michael Beyeler, Eduardo Fernandez. “Deep Learning–Based Control of Electrically Evoked Activity in Human Visual Cortex”. In Submission (Nature Communications)

Antonio Lozano*, **Jacob Granley***, Cristina Soto-Sanchez, Fabrizio Grani, Leili Soo, Al-

fonso Rodil, Rocio Lopez-Peco¹, Adrian Villamarin-Ortiz, Lily Turkstra, Michael Beyeler, Eduardo Fernandez. “Human-in-the-Loop Optimization reveals enhanced Temporal Patterns of Intracortical Microstimulation in the Blind Visual Cortex”. In Preparation

Jacob Granley*, Lily M Turkstra*, Hannah Stone, Fabrizio Grani, Leili Soo, Alfonso Rodil, Cristina Soto-Sanchez, Tommy Sprague, Eduardo Fernandez, Michael Beyeler. “Spiking Correlates of Working Memory in Blind Human Visual Cortex”. In preparation

Professional Experience

Machine Learning Research Fellow, LGS Labs (CACI) 2019–2020

Fully funded joint Masters program between CACI and Colorado School of Mines

- *Project Lead:* Developed a proof-of-concept machine learning system for real-time privacy attacks on IoT and embedded systems networks. One highlighted result was a novel tree-based algorithm that succeeded in zero-day device identification of encrypted 802.11 IoT network traffic when deployed in the wild. (Python)
- *Project Lead:* Designed an autoencoder-based anomaly detection algorithm to flag rogue or malicious cell towers, which correctly identified 99.8% of rogue cell towers based on transmission and system information blocks. Industry experts could not design a rogue tower that fooled the algorithm.

Software Engineering Intern, Lockheed Martin 2018

Feature development, bug fixes, and testing for a large-scale cross-domain security product.

Field Session Intern, The Zero Card 2017

Research and design of machine learning algorithms to classify and cluster medical claims data.

Abstract

Deep learning, perceptual modeling, and human-in-the-loop stimulus
optimization for artificial vision

by

Jacob Granley

Visual prostheses (‘bionic eyes’) aim to restore sight by electrically stimulating the retina or cortex, but current systems lack the intelligence to deliver consistent, high-quality percepts. This dissertation contributes to the development of a ‘Smart Bionic Eye’—a model-informed and user-adaptive vision restoration system—by introducing a computational framework that integrates deep learning, perceptual modeling, and human-in-the-loop optimization. The work begins with a data-driven model of phosphene appearance that predicts how perceptual features such as brightness, size, and shape vary with stimulus parameters. Trained on data spanning years of psychophysics and neuroanatomy, the model generalizes across electrodes and stimulation conditions and serves as the foundation for informed stimulus design. To solve the inverse problem of generating the electrical stimulus for a target percept, a deep neural network encoder is trained to invert the perceptual model. This encoder enables end-to-end optimization and consistently outperforms standard stimulation strategies in simulated users. To handle user variability and perceptual drift over time, the framework incorporates human-in-the-loop optimization using preferential Bayesian methods. This approach adapts stimulation strategies based on real-time user feedback and quickly converges to personalized solutions. Studies with sighted participants viewing simulated prosthetic vision demonstrate the method’s effectiveness and robustness to noise and model mismatch. Finally, the system

is extended to the human visual cortex. Using neural recordings from a blind participant implanted with a 96-channel Utah array, a deep model is trained to predict single-trial neural responses and synthesize stimulation patterns that evoke targeted activity. Both inverse networks and gradient-based controllers outperform conventional techniques and better modulate evoked activity. Together, these contributions establish a scalable framework for intelligent visual prostheses that adapt to individual users and bring the Smart Bionic Eye closer to clinical reality.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Aims and contributions	3
1.3	Organization	5
1.4	Permissions and Attributions	6
2	Background	7
2.1	Visual system and impairments	7
2.2	Introduction to visual prostheses	10
2.3	What do visual prosthesis users see?	11
2.4	Models of prosthetic vision	12
2.5	Stimulus optimization	16
3	Computational Model of Phosphene Appearance for Epiretinal Prostheses	22
3.1	Summary	22
3.2	Introduction	23
3.3	Methods	25
3.4	Evaluation	30
3.5	Discussion	33
4	Deep Neural Networks for Stimulus Optimization	35
4.1	Summary	35
4.2	Introduction	36
4.3	Background	38
4.4	Related Work	40
4.5	Methods	42
4.6	Results	48
4.7	Discussion	51
4.8	Conclusion	54
4.9	Acknowledgements	54

5	Human-in-the-Loop Optimization for Deep Stimulus Encoding	55
5.1	Summary	55
5.2	Introduction	56
5.3	Background and Related Work	59
5.4	Methods	62
5.5	Results	68
5.6	Discussion	74
5.7	Follow-up in Sighted Humans	76
5.8	Acknowledgements	86
6	Deep Learning-Based Control of Electrically Evoked Activity in Human Visual Cortex	87
6.1	Summary	87
6.2	Introduction	88
6.3	Results	91
6.4	Discussion	101
6.5	Methods	106
6.6	Data and Code Availability	120
6.7	Acknowledgments	124
7	Conclusion and Future Directions	125
A	Chapter 4 Appendix	128
A.1	Surrogate Model	128
A.2	Hyperparameter Selection	131
A.3	COCO Dataset	133
A.4	Predicted Stimuli	134
A.5	COCO Patient-to-Patient Variations	136
A.6	Modeling Other Patient-to-Patient Variations	138
A.7	Simulating Higher-Resolution Implants	139
A.8	Mis-Specified Patient-Specific Parameters	141
B	Chapter 5 Appendix	143
B.1	Phosphene Model	143
B.2	Deep Stimulus Inversion	149
B.3	Human-in-the-Loop Optimization	150
C	Chapter 6 Appendix	159
C.1	Supplementary Figures	160

Chapter 1

Introduction

1.1 Motivation

For millions of individuals living with sensory impairment, implanted neuroprostheses offer a promising route toward restoring lost sensory function by electrically stimulating neurons in brain areas responsible for perception. Devices targeting vision [1–3], audition [4], somatosensation [5, 6], alleviating motor pain [7], and treating neurological disorders [8] are in various stages of development. Yet all these neural interfaces face a fundamental challenge: interfacing with highly non-linear networks of biological neurons whose role in perception is not fully understood. Due to the limited spatial resolution of electrical stimulation, prostheses evoke neural response patterns that are unnatural to the brain, often resulting in percepts that are artificial, distorted, or variable [9–12].

A visual prosthesis electrically stimulates neurons along the visual pathway to restore visual perception [13], and is the only restorative treatment option for many sources of irreversible blindness [14]. However, increasing evidence now suggests that the vision provided by these devices differs substantially from natural eyesight [9, 15–19]. Despite their potential, visual prostheses rarely elicit percepts beyond simple shapes or flashes of

light. Phosphenes (i.e., percepts induced by electrical stimulation) are often distorted, difficult to interpret, and typically do not linearly combine to provide form vision. Prosthesis users often report infrequent device use, and struggle to find utility in the device for common, everyday tasks [20].

These challenges have motivated the development of a ‘smart bionic eye’, which focuses on restoring *useful* vision now rather than the currently infeasible *natural* vision [21]. In line with this, we identify two major limitations holding back prosthetic systems right now.

- 1 **Naive stimulation strategies.** There now exists decades of literature describing the various, complicated effects of stimulation on visual perceptions. However, current stimulation strategies do not properly leverage these effects when selecting electrical stimuli, instead assuming that each electrodes merely produce focal dots which nicely add up to construct a visual perception. In practice, this assumption leads to suboptimal stimuli which elicit distorted and unrecognizable perceptions.
- 2 **Variance in visual outcomes.** The phosphenes elicited from electrical stimulation vary dramatically, not only between different users, but also in the same user over time. This likely stems from a variety of biopsychosocial factors such as tissue-electrode interface, local neuroanatomy, plasticity, degeneration, the current underlying brain state, and other perceptual and cognitive processes. Current prostheses are not able to adapt to these variations, again producing suboptimal stimuli not tailored to an individuals perception.

1.2 Aims and contributions

The goal of this dissertation is to use techniques from computer science and deep learning to address these two challenges and aid in the development of a smart bionic eye. The first two studies presented in this thesis develop accurate computational models and use them to inform deep neural network-based stimulation strategies, addressing challenge 1. The later two studies then address challenge 2 by introducing a human-in-the-loop adaptive calibration procedure for deep learning-based stimulation strategies, and, as a proof-of-concept, deploy our deep learning framework to control neural activity in a human with an intracortical prosthesis.

Specifically, the contributions of this thesis are as follows:

- **Psychophysically validated phosphene models** for epiretinal prostheses. We first introduce a biologically inspired predictive model which is driven by data aggregated across 5 studies to accurately modulate phosphene size, brightness, and shape with amplitude, frequency, and pulse duration while being computationally efficient and differentiable. This evaluation, on data from two separate studies, also provides a benchmark for quantitative comparison. We later improve on this with a computationally-simpler model, which precisely modulates phosphene appearance as a function of both stimulus parameters and electrode location, representing the current state of the art in epiretinal phosphene prediction.
- **A deep-learning framework for stimulus encoding.** We introduce an encoder-decoder framework, where a deep neural network (*encoder*) is trained to invert a phosphene model (*decoder*), outputting stimuli which optimally reconstruct target images. This deep stimulus encoder demonstrates drastic improvements over naive strategies across a wide range of simulated patients. This framework is used extensively throughout the rest of the thesis, and similar approaches have been

prevalent in stimulus encoding research thereafter.

- **A deep human-in-the-loop optimization procedure** to quickly adapt stimulation strategies to individual users. In this procedure, Preferential Bayesian optimization is used to tune the deep stimulus encoder to produce percepts that are preferred by the user. We show this calibration leads to personalized, high-quality stimulation strategies for both simulated and sighted humans viewing simulated prosthetic vision, and is robust to noisy user feedback and misspecifications in the underlying phosphene model. This allows for stimulus encoders to be deployed on new patients requiring zero knowledge of their exact implant placement or phosphene perception.
- **First in human neural control via deep learning.** As a proof-of-concept, we deploy the encoder-decoder architecture in a human with an intracortical prosthesis to predict and shape neural recordings. We demonstrate this method can successfully drive neural activity toward target activation patterns, and can be conditioned on the brain’s resting state activity to provide better control. We also show that recorded neural activity predicts detection, brightness, and color more accurately than stimulation parameters alone, justifying control objectives at the level of neural population responses, and paving the way for future closed-loop stimulation strategies.

Together, these studies make significant progress towards enabling smarter stimulation strategies which are model-informed, are adaptive to individual users, and, eventually, incorporate closed-loop control of neural and perceptual responses. They also take the first steps towards deploying such systems in humans with real visual prostheses, a considerable technological hurdle. This will help aid in the development of smarter future visual prostheses, hopefully translating to more functional visual outcomes, offering an

effective treatment option to improve the lives of those living with blindness.

1.3 Organization

The remainder of this thesis is organized as follows.

Chapter 2 presents a relevant background on blindness, visual prostheses, computational modeling, and stimulus encoding, as well as a literature review on some of the established techniques in these areas.

Chapter 3 introduces a data-driven, biologically-inspired computational model of phosphene appearance for epiretinal prostheses. This model utilized aggregate data from five different source papers to constrain a phenomenological model which, for the first time in the literature, accurately modulated phosphene location, shape, size, brightness, orientation, and elongation with amplitude, frequency, pulse duration, and electrode location. Importantly, this model can be fit to individual patients using psychometric data, and was validated against data from two independent studies.

Chapter 4 presents a study introducing an end-to-end, encoder-decoder framework for stimulus optimization in visual prostheses. Here, a deep neural network stimulus encoder (DSE) was trained to invert the phosphene model of evoked visual perceptions, similar to an autoencoder. When evaluated in simulated patients, the deep stimulus encoder dramatically improves perception compared to conventional stimulation strategies.

Chapter 5 summarizes our work on making deep stimulus encoders (DSEs) adapt to individual users, since every prosthesis user experiences dramatically different artificial vision, often changing over time. To this end, we integrated human-in-the-loop Bayesian optimization (HILO) into the deep stimulus encoder, enabling it to adapt to individual prosthesis users. After a quick and easy calibration procedure, HILO learned personalized, high-quality stimulation strategies for simulated users, and was quite robust to noisy

user feedback and misspecifications in the underlying phosphene model. This chapter also introduces a new, state of the art phosphene model, and presents a brief case study, where the simulated results of HILO were confirmed in sighted humans viewing simulated prosthetic vision.

Chapter 6 presents our deployment of this framework to a human with an intracortical visual prosthesis. Since there is not yet an accurate phosphene model for intracortical prostheses, we instead utilized deep stimulus encoding to drive neural activity towards desired states. Two complimentary control strategies are introduced: a learned encoding network for real-time stimulation synthesis and a gradient-based optimizer for precise targeting of desired neural responses, both of which significantly outperformed existing mechanisms in shaping neural activity, and represent the first published methods for deep neural network-based activity shaping in human visual cortex. This chapter also presents deep models that decoded binary perception, brightness, and color from neural activity, underscoring the value of our neural population control framework.

Chapter 7 concludes this thesis, discussing implications of the main contributions, limitations, and potential future studies.

1.4 Permissions and Attributions

Chapters 3 and Section 5.7 are ©IEEE, as most of the content is derived from [22, 23] but under this copyright authors retain full rights to reuse their own material. Chapters 4 and 5 are primarily based on content from [24, 25], and authors retain full copyright. All other figures are self-owned or are works with an appropriately cited Creative Commons license. All text and ideas not considered common knowledge are cited appropriately with references provided in the Bibliography section.

Chapter 2

Background

2.1 Visual system and impairments

It is estimated that about 80% of sensory information is acquired through vision [26, 27], making it a critical part of human perception and cognition. Vision begins as light enters the eye through the cornea and lens, which focus the light onto the retina (See Figure 2.1 for an overview). On the very back of the retina, light activates rod and cone photoreceptors, which synapse onto bipolar cells, and, eventually, retinal ganglion cells (RGCs). Enough activation will cause the RGCs to generate action potentials. Axons from nearby RGCs form bundles and arc from cell bodies to the optic disc where they aggregate into the optic nerve (See Figure 3.2), carrying visual information to the brain. The optic nerves originating in each eye cross into the opposite hemisphere, where visual information is first preprocessed and relayed via the Lateral Geniculate Nucleus (LGN) of the thalamus before being directed to the primary visual cortex (V1).

After V1, visual input is processed in parallel through both the dorsal stream (parietal lobe), which processes information such as motion and spatial awareness, and the ventral stream (temporal lobe), which is responsible for recognizing different objects, faces, and

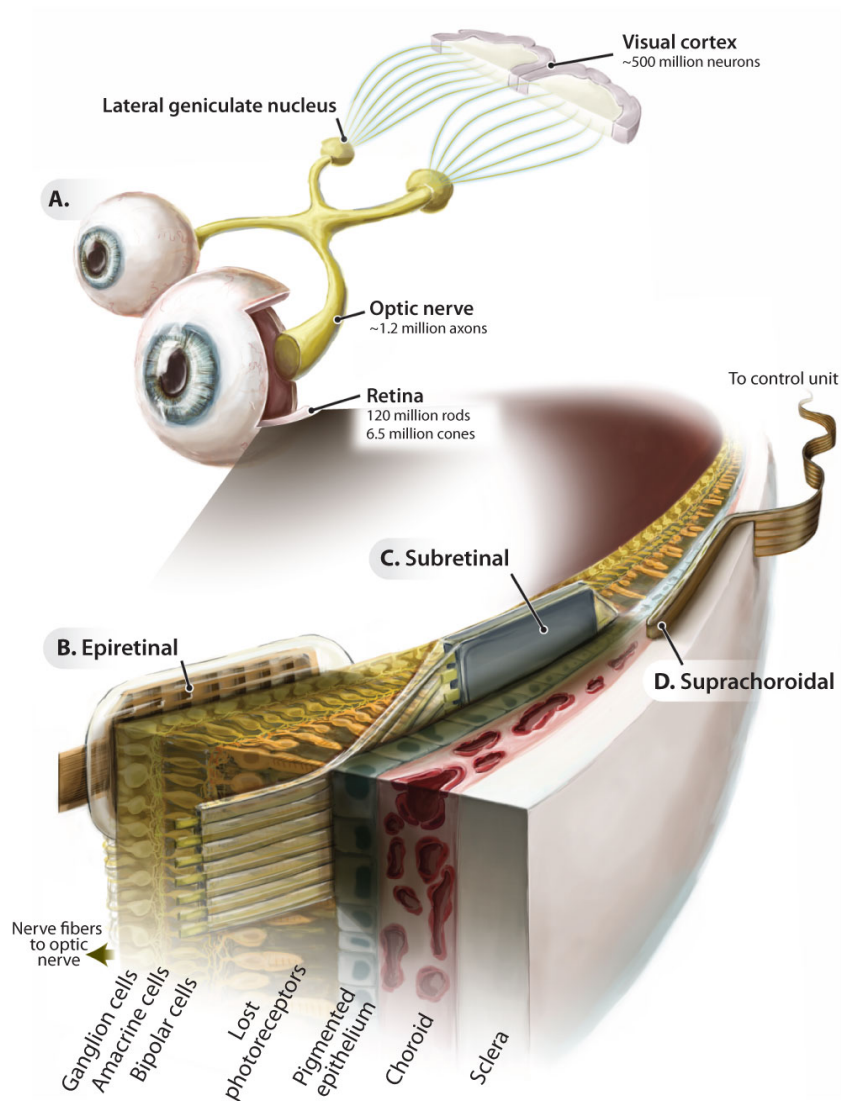


Figure 2.1: Overview of the visual pathway and common visual prosthesis implant locations. From Eberhart Zrenner. “Fighting Blindness with Microelectronics”. In: *Science Translational Medicine* 5.210 (Nov. 2013). Publisher: American Association for the Advancement of Science, 210ps16–210ps16. Reprinted with permission from AAAS.

processing color [29]. Generally, cells in lower visual areas (e.g. V1-V3) activate selectively for simple features in the image, such as lines or edges at certain orientations, and are organized retinotopically, whereas cells in higher visual areas are generally responsive to semantic content in the visual scene regardless of their position or exact appearance, such as to certain objects or people. In total, more than half of the human cortex processes or is modulated by visual input [27, 30].

Cause	% Affected	Pathway	Reversible?
Cataract	20%	Lens	Yes
Age-related Macular Degeneration (AMD)	16%	Retina	No
Glaucoma	14%	Optic nerve	No
Uncorrected refractive error	13%	Cornea/Lens	Yes
Retinitis Pigmentosa	6%	Retina	No
Diabetic Retinopathy	5%	Retina	If treated early
Corneal disease	3%	Cornea	Yes
Other	23%	Varies	Usually no

Table 2.1: Leading causes of total or profound blindness aggregated across developed regions as of 2015 [31, 32]. Percent affected is relative to the total blind population in those regions.

Damage to the visual system can result in conditions ranging from minor impairment to total blindness. As of 2020, there were 49 million people with total blindness [33], a number that is expected to rise to 115 million by 2050 [34]. Causes vary regionally: in developing countries most cases are reversible through cataract surgery and access to proper medical care and nutrition [35]. However, in developed and high-income regions, more than 60% of vision loss is irreversible (see Table 2.1). Many alternative treatment options are being developed for specific conditions, such as gene therapy [36], pharmacology [37], retinal sheet transplants [38], optogenetics [39], and stem cell therapy [40]. But for many conditions, the only restorative option is a visual prosthesis [14, 41], or, colloquially, a ‘Bionic Eye’.

2.2 Introduction to visual prostheses

Visual prostheses are a promising treatment option for millions of people living with irreversible vision loss [13, 42]. Visual prostheses use electrical stimulation via an implanted microelectrode array to directly excite visual neurons, effectively bypassing damaged parts of the visual system (See Figure 2.4, left). The brain then interprets this neural activity as visual perceptions, or ‘phosphenes’. Typically, visual surroundings are captured via a glasses-mounted camera and subsequently sent to a vision processing unit (VPU), which is responsible for image preprocessing and selecting a representative stimulus to be delivered via the implant. Due to safety constraints, stimuli are generally limited to charge-balanced square-wave biphasic pulse trains of varying amplitude, frequency, and pulse duration.

Since their beginnings over 50 years ago [43, 44], visual prostheses have been developed to electrically stimulate neurons in the retina [45–48], optic nerve [49], lateral geniculate nucleus [50], or early visual cortex [18, 19, 51, 52] (Figure 2.1). As technology has advanced, prostheses have progressed from crude, bulky laboratory devices with small electrode counts towards compact, wearable, bi-directional interfaces with the potential for modern video processing, AI integration, and up to thousands of electrodes. Currently, there are at least 55 different devices in various stages of development or deployment, and more than 500 people have received implants worldwide [53].

The two prostheses most relevant to the work in this thesis are Argus II [45] and Cortivis (Utah array [54]) [18]. Argus II is the most popular prosthesis, with over 350 recipients worldwide. In 2013, it became the first device in the US to obtain FDA approval for the treatment of retinitis pigmentosa [55], and was discontinued in 2019. It is an epiretinal device consisting of 60 $200\mu\text{m}$ electrodes in a 6×10 rectangular arrangement with $575\mu\text{m}$ pitch [56, 57] (spanning approximately 20 degrees visual angle). The Cortivis visual prosthesis project is an ongoing clinical trial currently recruiting their 5th participant

[58]. The device consists of a Utah array [54] implanted in V1 with 96 penetrating microelectrodes in a 10x10 grid with 400 μ m pitch (spanning approximately 2 degrees visual angle). In addition to stimulation, the Utah array is capable of simultaneously recording activity from the neurons underneath the array.

2.3 What do visual prosthesis users see?

Although they have promising potential, current visual prostheses provide only rudimentary vision restoration. Across devices, phosphenes typically appear artificial or distorted, limited primarily to simple, monochromatic (usually yellow, white, or blue, but occasionally green, gold, orange, purple, or ‘blacker than black’ [42]) blobs, lines, streaks, or flashes of light [18, 59]. In most prostheses, the locations of phosphenes are also dependent on eye movements [60], and quickly flicker or fade [61] even with continued stimulation.

Even more limiting is the variability. Phosphene appearance can vary dramatically, not only across individuals with the same implant, but even across different electrodes, or over time on the same electrode. Phosphene appearance also varies by implant type. Epiretinal prostheses often produce streaky or elongated phosphenes [17], and subretinal implants have produced smaller, more focal phosphenes [48]. Meanwhile, intracortical implants have elicited very small phosphenes, and to date have shown larger patient-to-patient and day-to-day variations in color and shape [18]. For more details, see *e.g.* Table 1 of [62].

One primary reason for these distortions is imperfect selectivity: electrical stimulation will indiscriminately activate neurons within a wide vicinity of the stimulating electrode. This ignores the underlying circuits through which visual information naturally flows, leading to artificial-looking percepts. However, another, often overlooked limitation is the inherent nonlinearities in visual circuitry.

A common misconception is that each electrode in the array can be thought of as a pixel in an image; to generate a complex visual experience, one then simply needs to turn on the right combination of pixels. There is now a growing body of evidence that this ‘scoreboard’ model of multielectrode stimulation is not accurate to patient-reported perceptions [17, 18, 52, 59, 63–65], presumably due to the nonlinearity of underlying neural circuitry [66, 67]. While pairs of electrodes can sometimes elicit two distinct phosphenes, they often merge into one separate phosphene, especially if the distance between electrodes is small [63, 64]. Perceptions from stimulation on three or more electrodes are much more variable and, although multielectrode stimulation can be used to perform certain tasks, it typically does not produce the perception of recognizable forms [59, 64]. This had recently led to researchers suggesting that the quality of prosthetic vision will likely be limited by neurophysiological organization, rather than engineering constraints [68].

These challenges have tremendously limited the utility of current visual prostheses; many users turn on the device rarely or infrequently, enjoying it more as a novelty than as an effective tool for everyday use [20]. In order to provide more useful vision, next generation prostheses will require not only improved hardware, but also a more detailed understanding of the visual percepts prostheses evoke, smarter stimulation strategies which can leverage this understanding to effectively manipulate percepts, and calibration procedures to adapt stimulation parameters to individual users.

2.4 Models of prosthetic vision

Development of accurate models which can predict the perceptions resulting from stimulation is critical, as it will directly enable *in silico* preclinical testing of more effective stimulation strategies, implant designs, and image preprocessing, and can be adapted to virtual reality to enable psychophysical testing with humans viewing simulated prosthetic

vision [69]. Models of phosphene appearance have appeared under many names in the literature, including phosphene model, simulated prosthetic vision (SPV), forward model, and phosphene decoder. Here we present a background on relevant models in the literature.

We identify two general types of models: ‘bottom-up’ biophysical models and ‘top-down’ phosphene models. Biophysical models (e.g. [70–72] attempt to recreate the neuronal circuitry and anatomy of the retina or cortex, and typically output firing rates of populations of neurons. Biophysical models have two main limitations. First, although neuronal firing rates are correlated with perception [73], this relationship is not well understood, and perception is typically affected by many other factors such as top-down attention. Second, detailed biophysical modeling is challenging due to the complexity and variability of retinal and cortical circuitry, especially in the presence of degeneration. Even if accurate, user-specific biophysical models could be obtained, the level of detail required often makes them too computationally expensive for many use cases. Thus, we here focus primarily on phosphene models.

In contrast, phosphene models are phenomenological, meaning that although they may be inspired by biological properties, they are mainly constrained by behavioral data, and predict visual perception directly from electrical stimulation [74]. Typically they consist of two, largely separate components: a topographic mapping, which predicts *where* phosphenes will be located in visual field based on electrode location and visuotopic maps, and a separate set of equations describing *what* the phosphene’s appearance will be (shape, size, color, brightness). Lacking biologically detail, phosphene models must make careful use of large, well-constructed datasets to ensure accurate and generalizable prediction of percepts. An overview of recent phosphene models is given in Table 2.2, (see [62] for an earlier review).

In the retina, cells typically respond to light originating from the corresponding portion of the vertically inverted visual field, with some small nonlinearities near the fovea [86,

Table 2.2: Summary of phosphene models for visual prostheses, excluding those presented in this thesis.

Paper	Implant Modality	Model type	Description	Psychophysically Validated	Stimulus Modulation
Thompson et al. (2003) [75]	N/A	Spatial	Fixed circular dot per electrode		
Scoreboard model [17]	N/A	Spatial	Simple Gaussian dots, variable size per patient		
Horsager et al. (2009) [76]	Epiretinal	Temporal	Leaky-integrator dynamics; persistence and desensitization	Thresholds	Amplitude Frequency
Nanduri et al. (2012) [77]	Epiretinal	Spatiotemporal	Extends [76] with current-spread & suprathreshold stimulation	Thresholds Brightness Size	Amplitude Frequency
Beyeler et al. (2019) [17]	Epiretinal	Spatial	Phosphenes elongated along axons	Shape	
Avraham et al. (2021) [78]	Epiretinal	Temporal	Basic persistence/fading; motion effects		
Kasowski (2023) [79]	Epiretinal	Temporal	Two-stage charge accumulation and brightness decay		
Hou et al. (2024) [80]	Epiretinal	Temporal	Piecewise spectral and exponential model	Fading	
Romeni et al 2024 [81]	Optic nerve	Spatial	Geometric model converting firing rates into visual perceptions		
de Ruyter van Steveninck et al. (2022) [82]	Cortical	Spatial	Differentiable phosphene simulator with on-off circular phosphenes		
Granley et al. (2022) [83]	Cortical	Spatial	Uses brain-like neural networks to obtain V1-like feature maps and predict percepts		
Wang et al. (2023) [84]	Cortical	Spatial	MRI-derived, subject-specific maps for location and size		
van der Grinten et al. (2024) [85]	Cortical	Spatiotemporal	Differentiable cortical simulator with temporal dynamics; real-time	Thresholds Fading [†]	Amplitude Frequency Pulse width
Fine & Boynton (2024) [68]	Cortical	Spatiotemporal	Orientation, ocular dominance, simple & complex cells, fading	Thresholds Size Brightness [†]	Amplitude Frequency Pulse width

Psychophysically validated: properties which are quantitatively validated against human psychophysics data ([†]: Validated, but dataset size is comparatively small).

Stimulus modulation: which stimulus parameters the percepts are modulated by (must also be validated to count).

87]. Horsager *et al.* [76] predicted perceptual thresholds by convolving simulated pulse trains with a cascade of linear filters and nonlinear processing steps. Nanduri *et al.* [77] extended this model to generalize to suprathreshold stimulation, and included certain stimulus parameters (see Section 5.5.1). However, these models are unable to explain many reported spatial effects, such as phosphene elongation.

To address this, Beyeler *et al.* [17] showed that the phosphene shape elicited by epiretinal implants could be predicted by the spatial activation pattern of axon nerve fiber

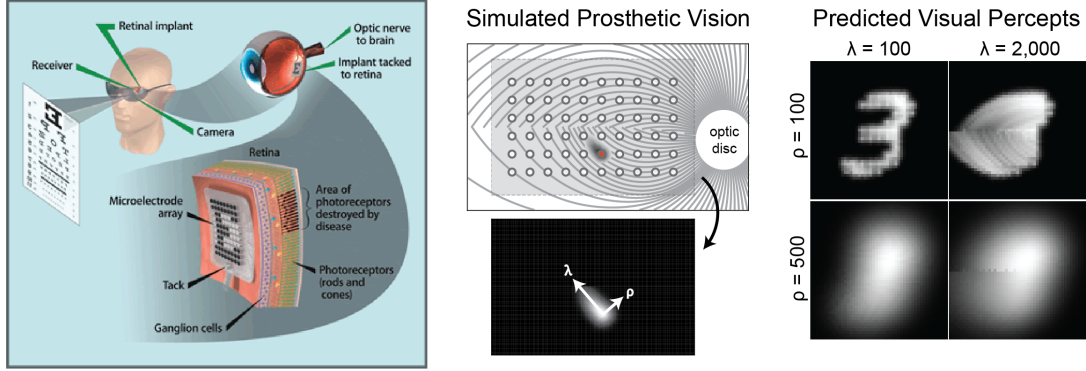


Figure 2.2: *Left:* Visual prosthesis. Incoming target images are transmitted from a camera to an implant in the retina, which encodes the image as an electrical stimulus pattern. Credit: U.S. Department of Energy (This image is considered public domain) *Center:* Electrical stimulation (red disc) of a nerve fiber bundle (gray lines) leads to elongated tissue activation (gray shaded region) and phosphenes whose shape can be described by two parameters, λ (axonal spread) and ρ (radial spread). *Right:* Predicted percepts for an MNIST digit using varying ρ and λ values.

bundles (Figure 2.4, middle). In this model, activation of a RGC cell body is assumed to decay exponentially as a function of (i) distance from the stimulating electrode, with decay rate ρ , and (ii) distance along the axon from the cell body, with decay rate λ , (Figure 2.4, right). A major limitation of all of these works is that although they predict aspects of spatial and temporal phosphene appearance, they cannot accurately predict the effects of changing stimulus parameters (e.g. amplitude, frequency, pulse duration) on phosphene appearance. In this thesis, we will present two new models which address these limitations.

In the cortex, phosphene models must account for cortical magnification, where the fovea takes up a disproportionately large amount of the cortical surface. All current models assume a wedge-dipole log-polar mapping [88]. Interestingly the much more accurate and patient-specific, anatomical MRI-derived mappings [89] have yet to be incorporated into any peer-reviewed phosphene model, however, Wang *et al.* did include MRI imaging to determine phosphene size and location [84]. The first cortical model to be incorporated into a differentiable simulated prosthetic vision pipeline was from de Ruyter van Stevenick

et al. [82], which assumed a simple on-off model of circular phosphenes. Improving on this, Van der Grinten *et al.* [85] included brightness and size modulation, in addition to temporal effects such as charge accumulation and phosphene fading. Finally, Fine *et al.*, [68] introduced a model which additionally incorporated orientation preference, and on-off simple and complex V1 cells. Predicting shape remains an outstanding challenge for cortical prostheses.

2.5 Stimulus optimization

Models are able to predict the phosphenes resulting from electrical stimulation. A smart visual prosthesis, however, must solve the inverse problem: what is the stimulus required to produce a target perception? A stimulus encoder (also referred to as an inverse model) is an algorithm responsible for selecting the optimal stimulus. Note that, although related, this process is distinct from image preprocessing (*e.g.*, scene simplification, segmentation, and edge enhancement [90]), which modify the captured visual scene into a more desirable, often simpler, target image.

The conventional encoding strategy employed by most prostheses simply sets the amplitude of each electrode proportional to the brightness of the corresponding pixel in the input image [45, 62]. Under this encoder, the stimulus is thus just a downsampled and rescaled version of the target image. This encoder works well if phosphenes on each electrode are small, non-overlapping shapes that combine linearly into a final percept, but fails in realistic settings where phosphenes are distorted and nonlinearly combine. This model is also referred to as ‘naive’, ‘direct mapping’, or ‘image downsampling’.

An large number of non-holistic stimulation strategies have been investigated [91], which attempt to improve an isolated response property or phosphene characteristic. These include current steering or local return electrodes to improve spatial resolution [92],

neural network optimization of stimulus parameters to achieve focal activation [93], and dynamic sequential stimulation to improve form recognition [52]. However, as these do not give a general algorithm on how to encode arbitrary images into stimuli, we do not consider them stimulus encoders and do not focus on them here.

An overview of stimulus encoders is presented in Table 2.3. In general, there are three categories of approaches (described below): dictionary learning, linear inversion, and deep learning-based inversion. Some encoders use a biomimetic approach [110], assuming that reproducing the brains natural neural code will elicit recognizable percepts. Others favor a perceptual approach, directly optimizing for human perception. Encoders also differ in whether they can be efficiently fit, or personalized to individual users (*Adaptable*), and in the evaluation platform with which they were tested (*e.g.*, in silico simulations, ex vivo, etc).

Dictionary approaches leverage the bidirectional capabilities of implants to record neural activity, building a dictionary of responses. Shah *et al.* [94, 95] used a greedy approach to iteratively select the stimuli that best recreate a desired neural activity pattern over a given temporal window, assuming that the brain integrates them into a coherent visual percept.

In contrast, the linear inversion approach assumes a forward model of the response to stimulation, and constructs a linear inverse to select stimuli. Spencer *et al.* [96, 111] proposed framing stimulus encoding as inversion of a forward linear-nonlinear model of neural activation patterns. Authors later extended this approach to allow for multipolar stimulation [111]. Fauvel *et al.* [97] used human in-the-loop Bayesian optimization to achieve encodings perceptually favored by the simulated patient, and use a linear inverse of the [17] model.

Since 2021, there has been a growing movement to frame stimulus optimization as an encoder-decoder deep learning problem. A model mapping stimulation parameters

to responses (neural activity or phosphenes) is used as the decoder. This model can be a DNN fit to data, or a fixed set of equations describing phosphene perception, such as those in Table 2.2. These are fixed, and a DSE neural network is trained as an inverse, to minimize the end-to-end reconstruction error between target and predicted phosphenes (for a more formal introduction, see Section 4.5). Thus, the stimulus encoder is model-aware, and learns to output stimuli which most faithfully reproduce the target image under the given phosphene model.

De Ruyter van Steveninck *et al.* [82] were among the first use this encoder-decoder architecture, with a simple on-off phosphene model. Since then a wide variety of similar approaches or improvements have been proposed, including targeting neural activity instead of phosphenes [112], surrogate networks instead of phosphene models [100, 112], more biologically plausible phosphene models [99], integration into a reinforcement learning virtual agent [101], safety constraints [102], classification-based reconstruction [103], improved inversion using conditional invertible DNNs [104], stroke-based ‘drawing’ of phosphenes to avoid axonal streaks [108], including a spiking recurrent network model of cortical activation [106], integrating eye movements [105], to reproduce downstream responses in optic nerve prostheses [98], adapted for learned image downsampling and validated in explanted mouse retinas [109], among others. Beyond visual prostheses, similar systems have been proposed for shaping activity in nonvisual areas in monkey [113], and optimizing stimuli in other sensory neuroprostheses [114, 115].

However, with the exception of [94, 95, 109], all of these visual prosthesis stimulus encoders have only been validated *in silico*, and have not yet been deployed in animal or human. There are a number of factors limiting practical applicability:

- 1 **Unrealistic phosphene decoders.** The majority of works do not use psychophysically validated phosphene models that can account for the effects of various stimulus

parameters on phosphene size, brightness, and, especially, shape. One reason for this is that extensively psychophysically validated phosphene models simply do not exist for many prostheses. However, even when they do exist, incorporating these highly nonlinear functions into a neural network and successfully training an encoder to invert them can be very difficult and unstable (see Chapter 4). Without accurate phosphene models, it is unlikely that the trained stimulus encoders will generalize to human participants.

- 2 **Cannot be calibrated for individual users.** Perceptions from visual prostheses vary dramatically user-to-user. Most phosphene models are fit to averages and cannot model individual variations, and thus will never be accurate for individual patients. For models that can, fitting parameters to new, unknown patients can often be infeasible, or require laborious collection of extensive psychophysics data. Next-generation models must increase adaptability in order to be effective in practical application (see Chapter 5).
- 3 **Limited data availability.** For many upcoming prostheses, there is simply not yet enough data to make accurate predictions of phosphene appearance. The collection and publication of large, openly-available psychometric and electrophysiological datasets by research groups leading studies will help mitigate this issue as more users are implanted.
- 4 **Difficulty of deployment.** The technological difficulties to integrate functional neural network models into real-world clinical or commercial systems in real time can make deployment infeasible, especially since the majority of current encoders are from academic groups with limited resources. Furthermore, deployment of a system that automatically selects stimulation parameters without manual oversight requires extensive precautions and clinical guidance ensuring safe stimulation parameters,

environment setups, and experimental design.

In this thesis, we take steps to address challenges 1, 2, and 4 by developing more accurate phosphene models, incorporating them into a neural network to predict optimal stimuli, adding a human-in-the-loop optimization (HILO) calibration procedure for individual users, and finally, with deployment to shape V1 neural activity in a human with an intracortical prosthesis. This provides the first steps towards a model-informed, personalized, closed-loop smart bionic eye that can hopefully provide more effective vision restoration to those living with profound blindness.

Table 2.3: Visual prosthesis stimulus encoders

Paper	Implant modality	Target type	Description	Validated decoder	Adaptable	Evaluation
Dictionary learning-based encoders						
Shah <i>et al.</i> (2019) [94]	Epiretinal	Hybrid	Greedy temporal dithering to match targets via RGC response dictionaries.	—	✗	Ex vivo (macaque)
Shah <i>et al.</i> (2024) [95]	Epiretinal	Hybrid	[94], stronger empirical evaluation of reconstruction vs static mapping.	—	✗	Ex vivo (macaque, rat)
Linear encoders						
Spencer <i>et al.</i> (2019) [96]	Epiretinal	Biomimetic	Algorithmic inversion of a linear-nonlinear neural decoder model.	✗	✗	In silico
Spencer <i>et al.</i> (2021) [96]	Epiretinal	Biomimetic	Multipolar activity shaping to increase spatial resolution	✗	✗	In silico
Fauvel <i>et al.</i> (2022) [97]	Epiretinal	Perceptual	Human-in-the-loop Bayesian optimization of encodings preferred by patients.	✗	✓	Sighted humans
Deep neural network-based encoders						
Romeni <i>et al.</i> (2021) [98]	Optic nerve	Hybrid	Genetic framework trained to invert a deep neural network (DNN) model of the ventral stream.	✗	✗	In silico
de Ruyter van Steveninck <i>et al.</i> (2022) [82]	Intracortical	Perceptual	End-to-end autoencoder with on-off phosphene decoder and secondary post-processing neural network.	✗	✗	In silico
van der Grinten <i>et al.</i> (2024) [99]	Intracortical	Perceptual	Extends [82] with increased biological realism, temporal optimization, and multiple stimulus parameters	✓ [†]	✗	In silico
Relic <i>et al.</i> (2022) [100]	Epiretinal	Perceptual	Deep encoder to invert a learned surrogate phosphene model.	✗	✗	In silico
Küçükoğlu <i>et al.</i> (2022) [101]	Intracortical	Perceptual	Task-oriented, end-to-end reinforcement learning in virtual environments.	✗	✗	In silico
Küçükoğlu <i>et al.</i> (2025) [102]	Intracortical	Perceptual	Introduces explicit safety constraints on stimulation.	✓ [†]	✗	In silico
Wu <i>et al.</i> (2023) [103]	Epiretinal	Perceptual	Classification-based reconstruction pipeline.	✗	✗	In silico
Wu <i>et al.</i> (2024) [104]	Epiretinal	Perceptual	Conditional invertible DNNs learn and invert a surrogate decoder.	✗	✗	In silico
Wu <i>et al.</i> (2024) [105]	Epiretinal	Perceptual	Gaze-contingent encoding.	✗	✗	In silico
Wang <i>et al.</i> (2023) [106]	Epiretinal	Hybrid	Projects predicted phosphenes to also model cortical activity	✓ [‡]	✗	In silico
Castro <i>et al.</i> (2024) [107]	Epiretinal	Biomimetic	Neural-activity shaping with a surrogate DNN fit to neural activity.	—	✓	In silico
Lavoie <i>et al.</i> (2024) [108]	Epiretinal	Perceptual	RL ‘stroke-based drawing’ to mitigate axonal streaks.	✗	✗	In silico
Leong <i>et al.</i> (2024) [109]	Epiretinal	Biomimetic	Actor encoder network to downsample images to evoke reliable neural responses, stimulation agnostic.	—	✗	Ex vivo (mouse)

Legend: ✓:Yes, ✗:No, —:not applicable or unnecessary.

Validated decoder: True only if the decoder model is psychophysically validated for phosphene shape and across all relevant stimulus parameters ([†]: validated for all except shape. [‡]: perceptual decoder validated, biomimetic not validated).

Adaptable: True if the encoder can efficiently adapt to individual users.

Chapter 3

Computational Model of Phosphene Appearance for Epiretinal Prostheses

3.1 Summary

Retinal neuroprostheses are the only FDA-approved treatment option for blinding degenerative diseases. A major outstanding challenge is to develop a computational model that can accurately predict the elicited visual percepts (phosphenes) across a wide range of electrical stimuli. Here we present a phenomenological model that predicts phosphene appearance as a function of stimulus amplitude, frequency, and pulse duration. The model uses a simulated map of nerve fiber bundles in the retina to produce phosphenes with accurate brightness, size, orientation, and elongation. We validate the model on psychophysical data from two independent studies, showing that it generalizes well to new data, even with different stimuli and on different electrodes. Whereas previous models

⁰This chapter contains work from the author’s previously accepted publication(s): (1) Jacob Granley and Michael Beyeler. “A Computational Model of Phosphene Appearance for Epiretinal Prostheses”. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. ISSN: 2694-0604. Nov. 2021, pp. 4477–4481

focused on either spatial or temporal aspects of the elicited phosphenes in isolation, we describe a more comprehensive approach that is able to account for many reported visual effects. The model is designed to be flexible and extensible, and can be fit to data from a specific user. Overall this work is an important first step towards predicting visual outcomes in retinal prosthesis users across a wide range of stimuli.

3.2 Introduction

Retinitis pigmentosa (RP) and age-related macular degeneration (AMD) are degenerative retinal diseases that lead to irreversible vision loss in more than 15 million people worldwide. As one promising treatment technology, retinal neuroprostheses [46, 116–118] aim to restore vision to these individuals by electrically stimulating the remaining retinal cells to evoke neural responses that are interpreted by the brain as visual percepts (“phosphenes”; see [119]).

However, a growing body of evidence suggests that the vision provided by these devices differs substantially from natural eyesight [9, 15–17]. Retinal implant users often report seeing distorted percepts and require extensive rehabilitative training to make use of their new vision [9]. Although single-electrode phosphenes are consistent from trial to trial, they vary across electrodes and users [17, 120]. Therefore, a deeper understanding of how electrical stimulation of the retina affects the quality of the generated artificial vision is crucial to designing more effective retinal implants.

A major outstanding challenge is to develop a computational model that can accurately predict visual outcomes for retinal implant users. Modeling the retinal response to electrical stimulation at a biophysical level (“bottom-up”) is challenging due to the complexity and variability of retinal circuitry in the presence of degeneration [121]. Even if an accurate user-specific biophysical model can be obtained, the detail required for simulation makes

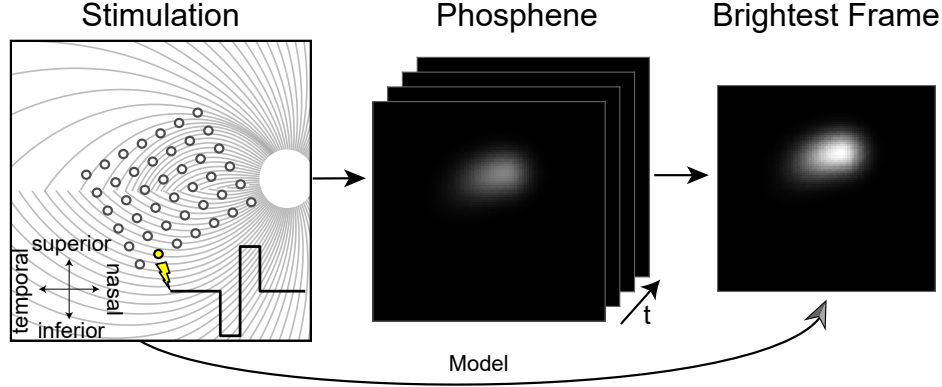


Figure 3.1: Phosphene model. A simulated biphasic pulse train is applied to a simulated epiretinal implant (*left*; for details see [74]), which evokes the percept of a flash of light whose brightness and shape evolve over time (*middle*). Rather than predicting the spatiotemporal properties of the elicited phosphene, the model directly predicts the brightest frame (*right*), which can be empirically validated using psychophysical data such as phosphene drawings and brightness ratings.

these methods too computationally expensive for many use cases [122].

In contrast, phosphene models are phenomenological (“top-down”) models constrained by behavioral data that predict visual perception directly from electrical stimuli [74]. To this end, Horsager *et al.* [76] predicted perceptual thresholds by convolving simulated pulse trains with a cascade of linear filters and nonlinear processing steps. Nanduri *et al.* [77] extended this model to generalize to suprathreshold stimulation. However, due to the number of free parameters and lack of an independent test set, these models should be viewed as descriptive rather than predictive models [77]. In addition, these models are unable to explain many reported spatial effects, such as phosphene elongation. To this end, Beyeler *et al.* [17] demonstrated that the phosphene shape elicited by epiretinal implants could be predicted by the spatial activation pattern of retinal nerve fiber bundles (NFBs). However, this cannot explain many reported temporal effects.

To address these challenges, we propose a phenomenological model constrained by both psychophysical [76, 77] and electrophysiological data [123] that predicts phosphene shape as a function of stimulus properties, such as amplitude, frequency, and pulse duration.

The model is designed to be flexible and extensible so that it provides good predictions on average but can also be fit to data from a specific user.

3.3 Methods

An overview of our model is given in Fig. 3.1. We assumed the subject is implanted with an epiretinal implant, such as Argus II [118] or POLYRETINA [124]. We focused on cathodic-first, square-wave, biphasic pulse trains, which make up the most common stimulus type in available devices. Given a stimulus, our model predicted the brightest “frame” of the percept seen by the user. Although the actual percept seen will likely grow and fade throughout the duration of stimulation, considering only the brightest frame made the problem tractable while allowing us to constrain the model with psychophysical data such as phosphene drawings and brightness ratings. A Python implementation based on `pulse2percept` [74] is available at <https://github.com/bionicvisionlab/2021-BiphasicAxonMap>.

3.3.1 Model Description

Our model extends the psychophysically validated axon map model [17] to account for a number of spatiotemporal effects. In this model, the shape of a phosphene generated by an epiretinal implant depended on the retinal location of the stimulating electrode. Because RGCs send their axons on highly stereotyped pathways to the optic nerve [125], an electrode that stimulates nearby axons would antidromically activate RGC bodies located peripheral to the point of stimulation, leading to percepts that appear elongated in the direction of the underlying NFB trajectory (Fig. 3.2, *left*). The model assumed that an axon’s sensitivity to electrical stimulation decayed exponentially as a function of (i) distance from the stimulating electrode, with decay rate ρ , and (ii) distance along

the axon from the cell body, with decay rate λ (Fig. 3.2, *right*). Since the maximum can occur either along the NFB or at the end of the NFB on the cell body, this model allows for both antidromic axonal and direct somatic RGC activation.

We extended this model with three new terms (F_{bright} , F_{size} , and F_{streak} ; described in detail below) that controlled how a percept's brightness, size, and streak length varied as a function of stimulus amplitude, frequency, and pulse duration. The output of the model was an intensity profile $I(r, \theta)$ that corresponded to the perceived brightness of a phosphene (polar coordinates centered over the fovea):

$$I(r, \theta) = \max_{p \in R(\theta)} \sum_{e \in E} F_{\text{bright}} \exp \left(\frac{-d_e^2}{2\rho^2 F_{\text{size}}} + \frac{-d_{\text{soma}}^2}{2\lambda^2 F_{\text{streak}}} \right) \quad (3.1)$$

where $R(\theta)$ was the path of the axon to the point (r, θ) , p was one individual point along the path, d_e was the Euclidean distance from p to the stimulating electrode, E was the set of all electrodes, and d_{soma} was the distance from p to the cell body along the axon,

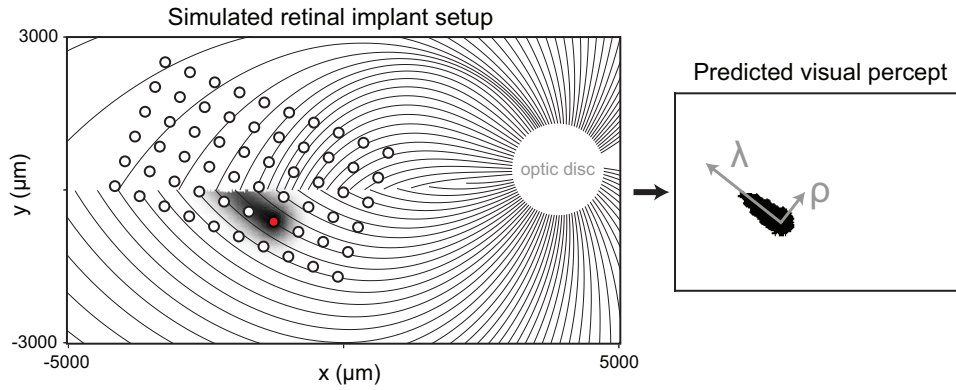


Figure 3.2: A simulated map of retinal NFBs (*left*) can account for visual percepts (*right*) elicited by retinal implants (reprinted with permission from [126]). *Left*: Electrical stimulation (red circle) of a NFB (black lines) could activate RGC bodies peripheral to the point of stimulation, leading to tissue activation (black shaded region) elongated along the NFB trajectory away from the optic disc (white circle). *Right*: The resulting visual percept appears elongated as well; its shape can be described by two parameters, λ (spatial extent along the NFB trajectory) and ρ (spatial extent perpendicular to the NFB). See Ref. [126] for more information.

given by the path integral over the NFB:

$$d_{soma} = \int_p^\theta \sqrt{R(\theta)^2 + \left(\frac{dR(\theta)}{d\theta}\right)^2} d\theta. \quad (3.2)$$

NFB paths ($R(\theta)$) were modeled as spirals originating at the optic disc and terminating at each ganglion cell body [125]. The spirals were fit using manual axon fiber tracings of fundus images of 55 human eyes (for details see [125]).

Finding functions for F_{bright} , F_{size} , and F_{streak} that accurately describe phosphene appearance reported by retinal implant users is a crucial component of the model. Ideally, these functions would be fit to perceptual data from a specific user, where F_{size} would modulate phosphene size as a function of stimulus parameters, F_{streak} would modulate phosphene elongation, and F_{bright} would modulate overall phosphene brightness (see Eq. 3.1). However, obtaining the amount of data needed to fit such a model can be challenging, and any user-specific fit would be unlikely to generalize to other individuals. An alternative is therefore to fit a general model to data averaged across users.

3.3.2 Model Fitting

Phosphene appearance is known to be affected by a number of stimulus parameters (Table 3.1). Both Horsager *et al.* [76] and Nanduri *et al.* [77] found a linear relationship between stimulus frequency and brightness as well as amplitude and brightness for Argus I. Nanduri *et al.* also found a linear relationship between amplitude and size, and found no evidence for a relationship between frequency and size. Weitz *et al.* [123] analyzed the effect of varying pulse duration using *in vitro* mouse retina, and found inverse relationships between pulse duration and streak length, as well as pulse duration and threshold. To the best of our knowledge, no data has been published measuring the effect of amplitude and frequency on perceived phosphene streak length.

Table 3.1: Stimulus parameters that affect phosphene appearance

	Amplitude		Frequency		Pulse Duration	
	Form	Source	Form	Source	Form	Source
Brightness	Linear	[76, 77]	Linear	[76, 77]	Power	[76]
Size	Linear	[77]	None	[77]	Power	[76]
Streak	Unknown	None	Unknown	None	Inverse power	[123]

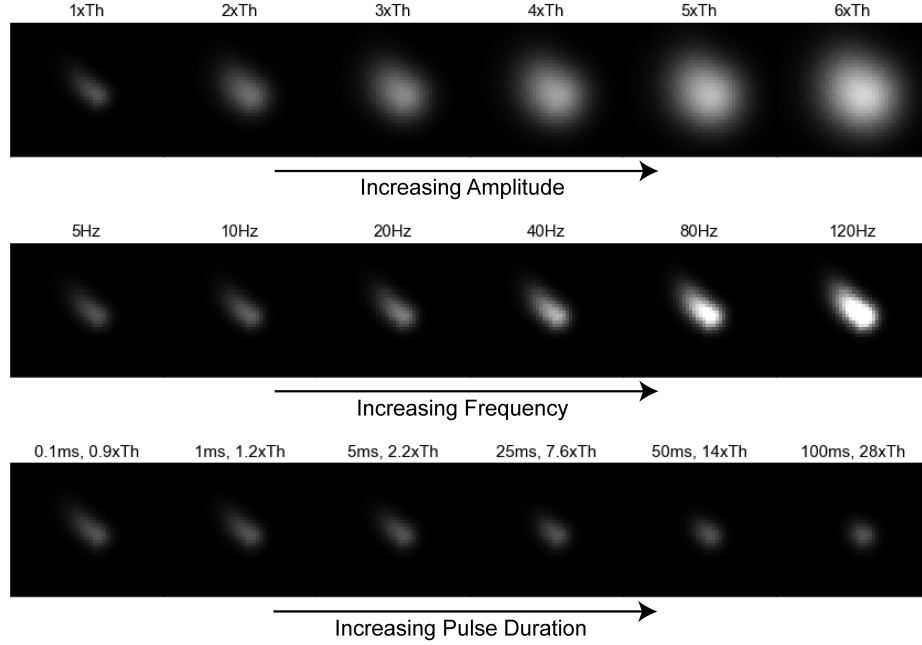


Figure 3.3: Predicted phosphene appearance as a function of amplitude (*top*), frequency (*middle*), and pulse duration (*bottom*). As pulse duration increased, we also decreased the amplitude to offset the decrease in threshold, isolating the effect of pulse duration. The stimulating electrode was a disk electrode with radius $200\text{ }\mu\text{m}$ located in the central superior retina. Stimulus parameters not shown were kept constant (amplitude: $1xTh$, frequency: 5 Hz , pulse duration: 0.45 ms , λ : $400\text{ }\mu\text{m}$, ρ : $200\text{ }\mu\text{m}$)

To fit the data from these three different studies, we first converted raw amplitude values to a factor a of the threshold current for each individual electrode. Threshold was defined as the amplitude necessary to produce a visible percept with 50 % probability for a reference stimulus with the same frequency and a pulse duration of 0.45 ms . Since increases in pulse duration corresponded with increases in threshold amplitude [123], we needed to account for the resulting indirect effect on size and brightness by scaling

amplitude a by a function of pulse duration, t :

$$\tilde{a}(t) = (A_0 t + A_1)^{-1} a. \quad (3.3)$$

The final model was given by:

$$F_{\text{bright}}(\tilde{a}, f) = A_2 \tilde{a} + A_3 f + A_4 \quad (3.4)$$

$$F_{\text{size}}(\tilde{a}) = A_5 \tilde{a} + A_6 \quad (3.5)$$

$$F_{\text{streak}}(t) = -A_7 t^{A_8} + A_9 \quad (3.6)$$

where $\tilde{a}$ was the scaled amplitude factor from Eq. 3.3, f was frequency in Hz, and t was pulse duration in ms. The scalars $A_0, \dots, A_9$ were the open parameters of the model (see code for values). All three equations were fit independently to data from the source papers using least squares regression.

Fig. 3.3 illustrates the model's predicted percepts for single-electrode stimulation as a function of stimulation parameters. Percepts reflected a number of phenomena reported by epiretinal implant users. First, streaks were elongated along the underlying NFB [17]. Second, stimulus amplitude modulated phosphene size and brightness, but frequency only modulated brightness [77]. Third, if pulse duration were to increase with other parameters held constant, the brightness and size would decrease due to the threshold increasing [123], causing the percept to quickly dim and fade. Therefore, we also adjusted amplitude (Fig. 3.3, *bottom*) to offset the change in threshold, illustrating that longer pulse durations lead to shorter streaks.

3.4 Evaluation

In order to conduct a quantitative evaluation, we compared model predictions against data from two independent studies with epiretinal prosthesis users.

3.4.1 Phosphene Appearance Across Stimulus Conditions

In the first experiment [77], an Argus I implant user was shown a reference stimulus as well as a test stimulus, which varied in either amplitude or frequency. The user was asked to rate the size or brightness of the test percept against the reference (Fig. 3.4, circles) on a scale from 0 to 100, with 10 meaning the brightness or size was equal to the reference, and 100 meaning 10x the reference.

To calculate predicted brightness, we ran both the reference and test stimuli through our model, and calculated the ratio between the predicted brightness of the test and reference pulse, multiplied by 10 to match the baseline in the psychophysical study. (Fig. 3.4A, B; solid black line). To estimate phosphene size, we counted the area with a predicted brightness greater than brightness at threshold, and again compared to the size of the reference stimulus (Fig. 3.4C, D; solid black line). Model predictions were compared to the baseline Nanduri model [77] (Fig. 3.4, dashed line).

Our model achieved a drastically better mean squared error (MSE) and R^2 than the Nanduri model for predicting brightness with amplitude modulation (MSE: 0.9 vs 11.1, R^2 : 0.91 vs -0.07) and frequency modulation (MSE: 2.1 vs 71.9, R^2 : 0.97 vs -0.19), and marginally better measures for predicting size with amplitude modulation (MSE: 0.16 vs 0.2, R^2 : 0.965 vs 0.957). Additionally, our model predicted elongated phosphenes that matched user drawings, following the axon map model described in [17], whereas the Nanduri model always predicted circular percepts.

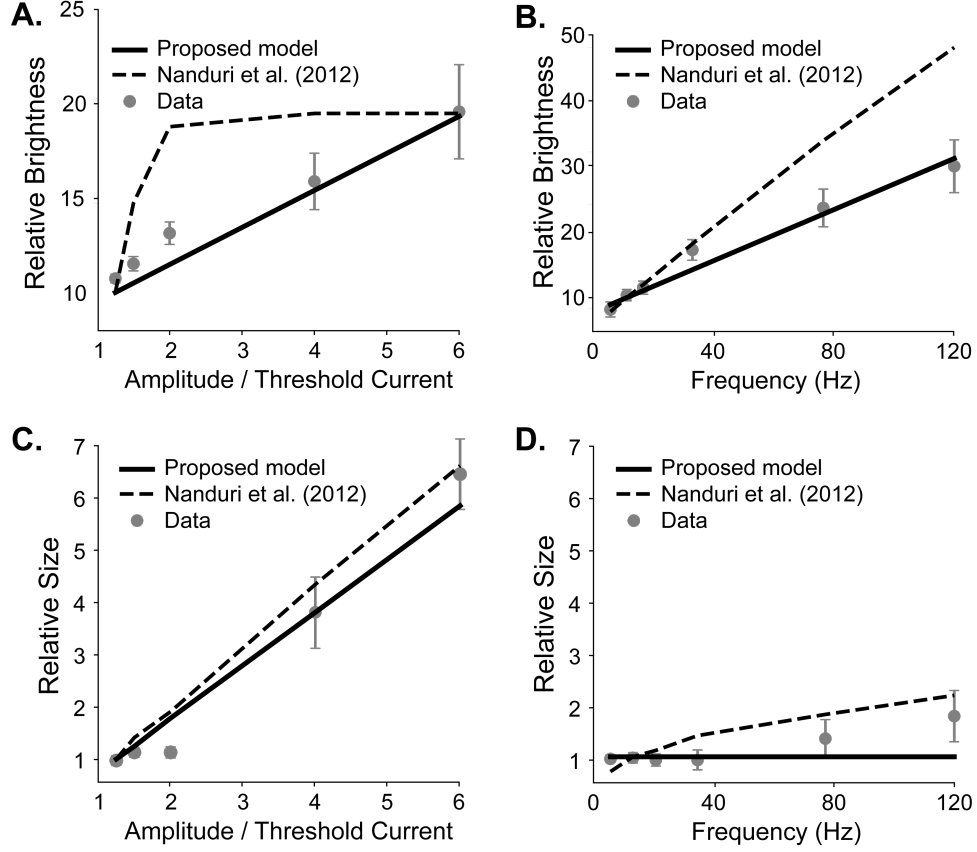


Figure 3.4: Brightness and size predictions for amplitude and frequency modulation on data from Nanduri *et al.* [77] (error bars: SEM). Solid lines: our model predictions. Dashed lines: Nanduri *et al.* model predictions.

3.4.2 Generalization to New Data

In order to evaluate the model’s ability to generalize, we tested on brightness rating data from Greenwald *et al.* [127] without refitting the model. The experiment conducted was a similar brightness rating task with two subjects, one being the same as in Nanduri *et al.* [77], but with different reference and test stimuli and electrodes.

The results are shown in Fig. 3.5. Here each data point is a single brightness rating on a particular electrode. The solid line shows our model’s predictions, the dashed line Nanduri *et al.* model predictions, while the dotted line depicts the linear model fit to the raw data using least squares. The Nanduri model performed poorly, showing it was

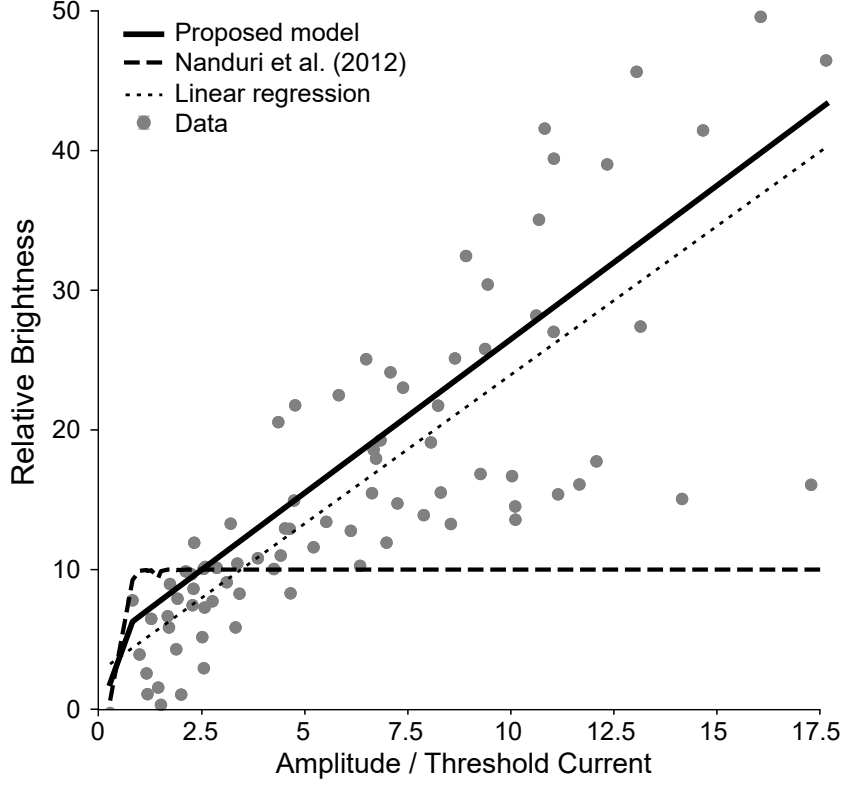


Figure 3.5: Brightness predictions for amplitude modulation on data from Greenwald *et al.* [127]. Solid line: our model prediction. Dashed line: Nanduri *et al.* predictions. Dotted line: Linear regression.

unable to generalize beyond the data presented in [77]. Our model performed significantly better, obtaining an MSE that was slightly higher than the optimal linear model (model: 49.5, linear regression: 44.5, Nanduri *et al.*: 160.9), and a slightly worse R^2 than the optimal linear model (model: 0.61, linear regression: 0.65, Nanduri *et al.*: -0.28).

The stimuli in this experiment used different pulse durations, amplitude ranges, electrodes, and one different subject than the Nanduri *et al.* study, yet the model was able to give accurate predictions, even with its pulse duration equations being fit to neurophysiological data from mouse retina. These preliminary results suggest that our model may generalize well to new data, even for novel stimuli on new electrodes or implant users.

3.5 Discussion

We introduced a computational model constrained by both psychophysical and electrophysiological data that predicts phosphene appearance in epiretinal prostheses as a function of stimulus amplitude, frequency, and pulse duration. Whereas previous models focused on either spatial or temporal aspects of the elicited phosphenes, this is the first model able to account for all the listed effects. The model’s ability to accurately predict phosphene appearance for the new user introduced in the Greenwald *et al.* experiment [127] demonstrates the robustness of our model, and suggests it may work fairly well for new users as is. Overall this work is an important first step towards predicting visual outcomes in retinal prosthesis users across a wide range of stimuli.

Furthermore, our model is open-source, highly modular, and can easily be extended to account for additional effects if data is available. For example, if data is available on how other implant or stimulus properties affect phosphene appearance (*e.g.*, stimulus waveform [128], electrode size [129], and electrode-retina distance [130]), Eqs. 3.3–3.6 could easily be updated to include the additional effects. As is, this work can be thought of as a general model that is able to make reasonable predictions on a patient-agnostic basis, similar in spirit to [76, 77]. However, since percepts vary widely across users, it is unlikely that any one model will be able to wholly describe every user’s unique perceptual experience. Future studies should thus focus on how this model could be fine-tuned to provide more accurate, patient-specific predictions.

In order to be used with a new user, Eqs. 3.3 - 3.6 should be refit to any brightness, size, and streak length ratings available from that user. Specifically, single-electrode phosphene drawings could be used to estimate patient-specific values for ρ and λ [17], and brightness ratings could be used to estimate F_{bright} [77]. However, little is known about how F_{size} (essentially a function of current spread in the tissue [77, 131]) and F_{streak}

(a function of axonal activation [70, 123]) vary from patient to patient. With more data, future studies could streamline this process and identify the smallest set of experiments needed to fit a patient-specific model.

Although the present model is able to predict phosphene appearance across various stimulus conditions, there are a few limitations that should be addressed in future work. First, although Eqs. 3.3–3.6 are relatively simple, there are still ten free parameters in the model, which is comparable to the model described in [77]. However, the generalization performance of our model is noticeably improved. Second, lack of data prohibited a more thorough evaluation of our model. The dataset used for generalization contains brightness measurements for single-electrode stimuli from two subjects—but did not include data for phosphene size or streak length, nor for multi-electrode stimuli. Future work should therefore investigate the model’s ability to generalize across multi-electrode stimuli as well as different users and devices.

Chapter 4

Deep Neural Networks for Stimulus Optimization

4.1 Summary

Sensory neuroprostheses are emerging as a promising technology to restore lost sensory function or augment human capabilities. However, sensations elicited by current devices often appear artificial and distorted. Although current models can predict the neural or perceptual response to an electrical stimulus, an optimal stimulation strategy solves the inverse problem: what is the required stimulus to produce a desired response? Here, we frame this as an end-to-end optimization problem, where a deep neural network stimulus encoder is trained to invert a known and fixed forward model that approximates the underlying biological system. As a proof of concept, we demonstrate the effectiveness of this hybrid neural autoencoder (HNA) in visual neuroprostheses. We find that HNA produces high-fidelity patient-specific stimuli representing handwritten digits and segmented images

⁰This chapter contains work from the author's previously accepted publication(s):
(1) **granley_hybrid_2022-1**

of everyday objects, and significantly outperforms conventional encoding strategies across all simulated patients. Overall, this is an important step towards the long-standing challenge of restoring high-quality vision to people living with incurable blindness and may prove a promising solution for a variety of neuroprosthetic technologies.

4.2 Introduction

Sensory neuroprostheses are emerging as a promising technology to restore lost sensory function or augment human capacities [14, 132]. In such devices, diminished sensory modalities (e.g., hearing [133], vision [18, 45], cutaneous touch [134]) are re-enacted through streams of artificial input to the nervous system. For example, visual neuroprostheses electrically stimulate neurons in the early visual system to elicit neuronal responses that the brain interprets as visual percepts. Such devices have the potential to restore a rudimentary form of vision to millions of people living with incurable blindness.

However, all of these technologies face the challenge of interfacing with a highly nonlinear system of biological neurons whose role in perception is not fully understood. Due to the limited resolution of electrical stimulation, prostheses often create neural response patterns foreign to the brain. Consequently, sensations elicited by current sensory neuroprostheses often appear artificial and distorted [9, 10]. A major outstanding challenge is thus to identify a stimulus encoding that leads to perceptually intelligible sensations. Often there exists a forward model, f (Fig. 4.1A), constrained by empirical data, that can predict a neuronal or (ideally) perceptual response to the applied stimulus (see [135] for a recent review). To find the stimulus that will elicit a desired response, one essentially needs to find the inverse of the forward model, f^{-1} . However, realistic forward models are rarely analytically invertible, making this a challenging open problem for neuroprostheses.

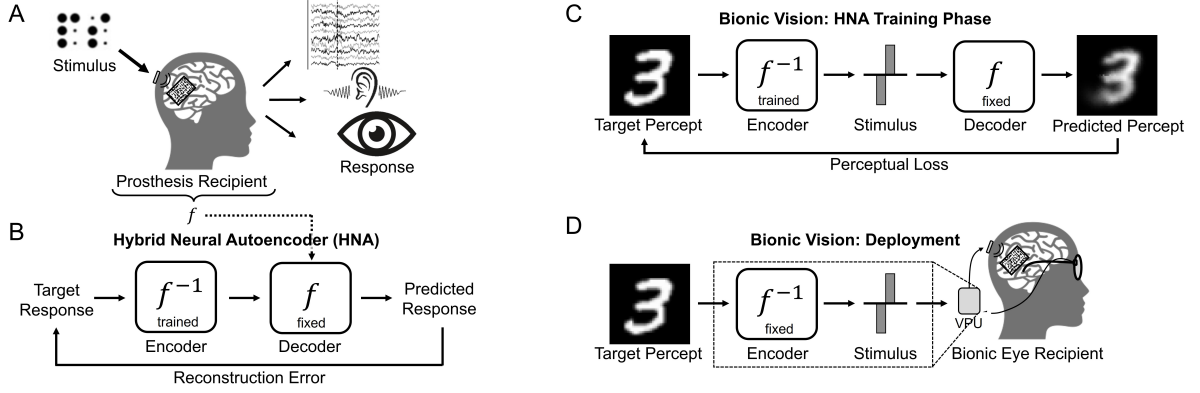


Figure 4.1: *A)* Sensory neuroprosthesis. A forward model (f) is used to approximate the neuronal or, ideally, perceptual response to electrical stimuli. *B)* Hybrid neural autoencoder (HNA). A deep neural encoder (f^{-1}) is trained to predict the patterns of electrical stimulation that elicit responses closest to the target. *C)* Visual neuroprostheses are one prominent application of HNA, where an encoder can be trained to predict the electrical stimulation needed to elicit a desired visual percept. *D)* The trained encoder is deployed on a vision processing unit (VPU), predicting stimuli in real-time that are decoded by the patient’s visual cortex.

Here we propose to approach this as an end-to-end optimization problem, where a DNN (*encoder*) is trained to invert a known, fixed forward model (*decoder*, Fig. 4.1B). The encoder is trained to predict the patterns of electrical stimulation patterns that elicit perception (*e.g.*, vision, audition) or neural responses (*e.g.*, firing rates) closest to the target. This hybrid neural autoencoder (HNA) could in theory be used to optimize stimuli for any open-loop neuroprosthesis with a known forward model that approximates the underlying biological system.

In order to optimize end-to-end, the forward model must be differentiable and computationally efficient. When this is not the case, an alternative approach is to train a surrogate neural network, $\hat{f}$, to approximate the forward model [100, 136–138]. However, even well-trained surrogates may result in errors when included in our end-to-end framework, due to the encoders’ ability to learn to exploit the surrogate model. We therefore also evaluate whether a surrogate approach to HNA is suitable.

To this end, we make the following contributions:

- We propose a hybrid neural autoencoder (HNA) consisting of a deep neural encoder trained to invert a fixed, numerical or symbolic forward model (decoder) to optimize stimulus selection. Our framework is general and addresses an important challenge with sensory neuroprostheses.
- As a proof of concept, we demonstrate the utility of HNA for visual neuroprostheses, where we predict electrode activation patterns required to produce a desired visual percept. We show that the HNA is able to produce high-fidelity, patient-specific stimuli representing handwritten digits and segmented images of everyday objects, drastically outperforming conventional approaches.
- We evaluate replacing a computationally expensive or nondifferentiable forward model with a surrogate, highlighting benefits and potential dangers of popular surrogate techniques.

4.3 Background

Sensory Neuroprostheses Recent advances in neural understanding, wearable electronics, and biocompatible materials have accelerated the development of sensory neuroprostheses to restore perceptual function to people with impaired sensation. Use of neuroprostheses typically requires invasive implants that elicit neural responses via electrical, magnetic, or optogenetic stimulation. Two of the most promising applications are cochlear implants, which stimulate the auditory nerve to elicit sounds [133], and visual implants (see next subsection) to restore vision to the blind. However, a variety of other devices are in development; for instance, to restore touch [134, 139] or motor function [140]. The latter differ from other sensory neuroprostheses in that they generate stimuli using motor feedback (*closed loop*) [141, 142]. In the absence of feedback (*open loop*), a proper stimulus encoding is paramount to the success of these devices.

Restoring Vision to the Blind For millions of people who are living with incurable blindness, a visual prostheses (*bionic eye*, Fig. 2.4, *left*) may be the only treatment option [119]. Analogous to cochlear implants, these devices electrically stimulate surviving cells in the visual pathway to evoke visual percepts (*phosphenes*), which can support simple behavioral tasks [45, 143, 144].

A common misconception is that each electrode in the array can be thought of as a pixel in an image; to generate a complex visual experience, one then simply needs to turn on the right combination of pixels [44]. However, recent evidence suggests that phosphenes often appear distorted (*e.g.*, as lines, wedges, and blobs) and vary drastically across subjects and electrodes [9, 18].

Phosphene appearance has been best characterized in epiretinal implants, where inadvertent activation of NFBs in the optic fiber layer of the retina leads to elongated phosphenes [17, 145] (Fig. 2.4, *center*). To this end, Granley *et. al* [22] developed a forward model to predict phosphene shape as a function of both neuroanatomical parameters (*i.e.*, location of the stimulating electrode) and stimulus parameters (*i.e.*, pulse frequency, amplitude, and duration). With this model, phosphenes are primarily characterized by two main parameters, ρ and λ , which dictate the size and elongation of the elicited phosphene, respectively (Fig. 2.4, *right*). These parameters can be determined using psychophysical tasks (*e.g.*, drawings, brightness ratings) [17, 22], and although they vary drastically across patients [17], they do not change much over time [16, 120]. Stimulation from multiple electrodes is nonlinearly integrated into a combined perception, and if two electrodes happen to activate the same NFB, they might not generate two distinct phosphenes.

4.4 Related Work

The conventional ‘naive’ encoding strategy sets the amplitude of each electrode to the brightness of the corresponding pixel in the target image [45, 62], making the stimulus a down-sampled version of the target. Although simple, this strategy only works with near-linear forward models, cannot account for real phosphene data, and often leads to unrecognizable percepts (Fig. 2.4, *right*) [9, 17].

Many alternative stimulation strategies have been proposed [91]. Shah *et al.* [94] used a greedy approach to iteratively select the stimuli that best recreate a desired neural activity pattern over a given temporal window, assuming that the brain would integrate them into a coherent visual percept. Ghaffari *et al.* [93] used a neural network surrogate model combined with an interior point algorithm to optimize for localized, circular neural activation patterns for individual electrodes. Fauvel *et al.* [97] used human in-the-loop Bayesian optimization to achieve encodings perceptually favored by the simulated patient. Spencer *et al.* [96] proposed framing stimulus encoding as inversion of a forward model of neural activation patterns, but to approximate the inverse, their approach either requires simplification or is NP-hard [96].

Van Steveninck *et al.* [82] proposed an end-to-end optimization strategy with a fixed phosphene model, similar to HNA. However, their approach crucially differs from ours in its inclusion of a secondary DNN to post-process the predicted phosphenes. This is a critical limitation, because a low reconstruction loss does not reveal whether a high-fidelity encoder was learned or whether the secondary decoder network simply learned to correct for the encoder’s mistakes. In addition, they used an unrealistic phosphene model that simply upscales and smooths a binary stimulus pattern. It is therefore not clear whether their results would generalize to real visual prosthesis patients.

Relic *et al.* [100] also utilized the end-to-end approach, but without the secondary

decoder network used in [82]. They used a more realistic phosphene model, which accounts for some spatial distortions (*e.g.*, axonal streaks), but not the effects of stimulus parameters. Since including a realistic phosphene model in the loop is not straightforward, they instead trained a surrogate neural network to approximate the forward model. We re-implemented Relic’s surrogate approach in this paper as a baseline method to compare against, as described in Section 4.5.

Taken together, we identified three main limitations of previous work that this study aims to address:

- 1) **Unrealistic forward models.** Most previous approaches (*e.g.*, [82, 94, 96]) use an overly simplified forward model that cannot account for empirical data [9, 17]. We overcome this limitation by developing (and inverting) a differentiable version of a neurophysiologically informed and psychophysically validated phosphene model [22] that can account for the effects of stimulus amplitude, frequency, and pulse duration on phosphene appearance.
- 2) **Optimization of neural responses.** Most previous approaches (*e.g.*, [94, 96]) focus on optimizing neural activation patterns in the retina in response to electrical stimulation (“bottom-up”). However, the visual system undergoes extensive remodeling during blinding diseases such as retinitis pigmentosa [146]. Thus the link between neural activity and visual perception is unclear. We overcome this limitation by inverting a phenomenological (“top-down”) model constrained by behavioral data that predicts visual perception directly from electrical stimuli [17, 22].
- 3) **Objective function** Most previous approaches rely on minimizing MSE between the target and reconstructed image. Although simple and efficient, MSE is known to be a poor measure of perceptual dissimilarity for images [147] and does not align well with human assessments of image quality [148]. We overcome this limitation by

proposing a joint perceptual metric that combines mean absolute error (MAE), VGG, and Laplacian smoothing losses.

4.5 Methods

Problem Formulation We consider a system where there is some known forward process f mapping stimuli to responses $f : \mathcal{S} \mapsto \mathcal{R}, f(\mathcal{S}) \subset \mathcal{R}$. In the case of visual prostheses, f may map stimuli to visual percepts. f may optionally be parameterized by patient-specific parameters ϕ .

Finding the best stimulus for an arbitrary target response $\mathbf{t} \in \mathcal{R}$ is equivalent to finding the inverse of f . However, since not every response can be perfectly reproduced by a stimulus, the true inverse of f is not well defined. We therefore seek the pseudoinverse (still denoted as f^{-1} for simplicity) instead, which outputs the stimuli that produce the closest response to $\mathbf{t}$ under some distance metric d :

$$f^{-1}(\mathbf{t}, \phi) := \arg \min_{\mathbf{s} \in \mathcal{S}} d(f(\mathbf{s}; \phi), \mathbf{t}). \quad (4.1)$$

This problem is straightforward to solve using an autoencoder approach, where a learned encoder f^{-1} is trained to invert the fixed decoder f (i.e., forward model).

Encoder We approximate the pseudoinverse f^{-1} with a DNN encoder $\hat{f}^{-1}(\mathbf{t}, \phi; \theta)$ with weights θ , trained to minimize the distance d between the target image $\mathbf{t}$ and predicted image $\hat{\mathbf{t}}$:

$$\min_{\theta, \phi \sim p(\phi)} d(\mathbf{t}, \hat{\mathbf{t}}) \quad (4.2)$$

$$\hat{\mathbf{t}} = f(\hat{f}^{-1}(\mathbf{t}, \phi; \theta); \phi), \quad (4.3)$$

where ϕ is sampled from a uniform random distribution spanning the empirically observed range of patient-specific parameters [17, 22].

We use a simple architecture consisting solely of fully connected (FC) and batch normalization (BN) [149] layers (1.4M trainable parameters). First, the target $\mathbf{t}$ is flattened and input to a FC layer. In parallel, the patient parameters ϕ are input to a BN layer and two hidden FC layers. Then, the outputs of these two paths are concatenated, and the combined vector fed through two FC layers, producing an intermediate representation $\mathbf{x}$. Amplitudes are predicted from $\mathbf{x}$ with a FC layer. The amplitudes are then concatenated with $\mathbf{x}$, put through a BN layer, and used to predict frequency and pulse duration, each with a FC layer. The outputs are merged into a stimulus matrix $\hat{\mathbf{s}}$. All layers use leaky ReLU activation, except for stimulus outputs, which use ReLU to enforce nonnegativity.

Decoder The HNA decoder is a differentiable approximation of the underlying biological system, and describes the transform from stimulus to response. For our decoder f , we use a reformulated but equivalent version of the model described in [22]. This model is specific to epiretinal prostheses; analogous models exist for other neuroprostheses (*e.g.*, auditory [150–155], tactile and somatosensory [156–160]), and could potentially be adapted for use with HNA. We use a square 15×15 array of $150\mu\text{m}$ electrodes, spaced $400\mu\text{m}$ apart and centered on the fovea. The size and scale of this device are motivated by similar designs in real epiretinal implants.

f takes as input a stimulus matrix $\mathbf{s} \in \mathbb{R}_{\geq 0}^{n_e \times 3}$, where the stimulus on each electrode ($\mathbf{s}_e$) is a biphasic pulse train described by its frequency, amplitude, and pulse duration. A simulated map of retinal NFBs is used to predict phosphene shape. Following [17], each retinal ganglion cells’ activation is assumed to be the maximum stimulation intensity along its axon. Axon sensitivity is assumed to decay exponentially with i) distance to the stimulating electrode (radial decay rate, ρ) and distance to the soma along the

curved axon (axonal decay rate, λ). To account for stimulus properties [22], ρ , λ , and the per-electrode brightness are scaled by three simple equations $F_{\text{size}}(\mathbf{s}_e, \phi)$, $F_{\text{streak}}(\mathbf{s}_e, \phi)$, and $F_{\text{bright}}(\mathbf{s}_e, \phi)$, respectively.

The brightness of a pixel located at the point $\mathbf{x} \in \mathbb{R}^2$ in the output image is given by

$$f(\mathbf{s}; \phi) = \max_{\mathbf{a} \in A} \sum_{e \in E} F_{\text{bright}}(\mathbf{s}_e, \phi) \exp \left(\frac{-\|\mathbf{x} - \mathbf{e}\|_2^2}{2\rho^2 F_{\text{size}}(\mathbf{s}_e, \phi)} + \frac{-d_s(\mathbf{x}, \mathbf{a})^2}{2\lambda^2 F_{\text{streak}}(\mathbf{s}_e, \phi)} \right) \quad (4.4)$$

where A is the cells' axon trajectory, E is the set of electrodes, $\phi = \{\rho, \lambda, \dots\}$ is a set of 12 patient-specific parameters, and d_s is the path length along the axon trajectory [125] from $\mathbf{a}$ to $\mathbf{x}$:

$$d_s(\mathbf{x}, \mathbf{a}) = \int_{\mathbf{a}}^{\mathbf{x}} \sqrt{A(\theta)^2 + \left(\frac{dA(\theta)}{d\theta} \right)^2} d\theta. \quad (4.5)$$

This model (f) can be fit to individual patients; however, it is computationally slow and not differentiable. For more details on these equations, see [22]. We therefore considered two approaches:

- **Differentiable Model:** We reformulated equations 4.4 and 4.5 into an equivalent set of parallelized matrix operations, implemented in Tensorflow [161]. Significant efforts were put towards developing a model optimized for XLA engines on GPU, resulting in speedups of up to 5000x compared to the model as presented in [22], enabling large-scale gradient descent. To enforce differentiability, we numerically approximated d_s using $|A| = 500$ axon segments per axon.
- **Surrogate Model:** We also implemented the surrogate approach from [100] as a baseline method, where f is approximated with another DNN $\hat{f}_\phi(\mathbf{s}; \theta_f)$ with weights θ_f . To achieve this we generated 50,000 percepts using randomly selected stimuli passed through f and fit a DNN to produce identical images. As f is highly dependent on patient-specific parameters ϕ , we generated new data and fit a separate surrogate model

for each ϕ in our experimental set. Specific implementation details of the surrogate are presented in Appendix A.1. Our implementation improves upon [100] by using the more advanced phosphene model described above, which accounts for effects of stimulus properties and allows for optimization of stimulus frequency in addition to amplitude.

Metrics To measure perceptual similarity, we use a joint perceptual objective consisting of a VGG [162] similarity term, a mean absolute error (MAE) term, and a smoothness regularization term. The MAE term is given by $L_{\text{MAE}} = \frac{1}{|\mathbf{t}|} \|\mathbf{t} - \hat{\mathbf{t}}\|_1$.

The VGG term aims to capture higher-level differences between images [82, 163]. The target image and reconstructed phosphene are input to VGG-19 pretrained on ImageNet [164], and the MSE between the activations on a downstream convolutional layer is computed. Let V_l be a function that extracts the activations of the l -th convolutional layer for a given image. The VGG loss is then defined as $L_{\text{VGG}} = \frac{1}{|\mathbf{t}|} \|V_l(\mathbf{t}) - V_l(\hat{\mathbf{t}})\|_2^2$.

We also include a smoothing regularization term. This term imposes a loss on the second spatial derivative of the predicted image. A low second derivative implies that where the target image does change, it changes slowly. We found this encouraged smoother, more connected phosphenes. To approximate the second derivative, we convolve the image with a Laplacian filter [165] of size k , denoted by $Lap(\cdot, k)$, and compute the mean. The smoothness loss is given by:

$$L_{\text{Smooth}} = \frac{1}{|\hat{\mathbf{t}}|} \sum_i \left(\frac{\partial^2}{\partial x^2} \hat{\mathbf{t}} \right)_i = \frac{1}{|\hat{\mathbf{t}}|} \sum_i Lap(\hat{\mathbf{t}}, k)_i. \quad (4.6)$$

Our final objective is the weighted sum of the three individual losses, given by Eq. 4.7,

where α and β are hyperparameters controlling the relative weighting of the three terms.

$$d = L_{\text{MAE}} + \alpha L_{\text{Smooth}} + \beta L_{\text{VGG}}. \quad (4.7)$$

We also implement a secondary metric to quantify phosphene recognizability, applicable only for the MNIST reconstruction task. We first pre-train a classifier network on the MNIST targets until it reaches 99% test accuracy, and then fix the weights. The relative accuracy (RA) is then defined as the ratio of the classifiers accuracy on the reconstructed images to its accuracy on the targets $RA = ACC/ACC(\mathbf{t})$. A perfect encoder would result in $RA = 100\%$. A similar process was not possible for the COCO task due to the possibility of having multiple objects in each target image.

Training/Optimization We trained using Tensorflow 2.7 [161] on a single NVIDIA RTX 3090 with 24GB memory. Stochastic gradient descent with Nesterov momentum was used to minimize the joint perceptual loss. We used a batch size of 16 due to memory limitations imposed by f . The amplitude, frequency predictions are scaled by 2, 20 respectively, while the pulse duration predictions were shifted by 1e-3 prior to being fed through the decoder. This encourages the initial predictions of the network to be in a reasonable range. The Laplacian filter size k is set to 5. We choose l to be first convolutional layer in the last block using cross validation (see Appendix A.2). Similarly, we perform cross validation to find the best values for α and β . Instead of using one value, we found that incrementally increasing the weighting of the VGG loss (β) from 0 while simultaneously decreasing the initially high weight on the smoothing constraint (α) was crucial for performance, especially when the range of allowed ϕ values was large (see Appendix A.2).

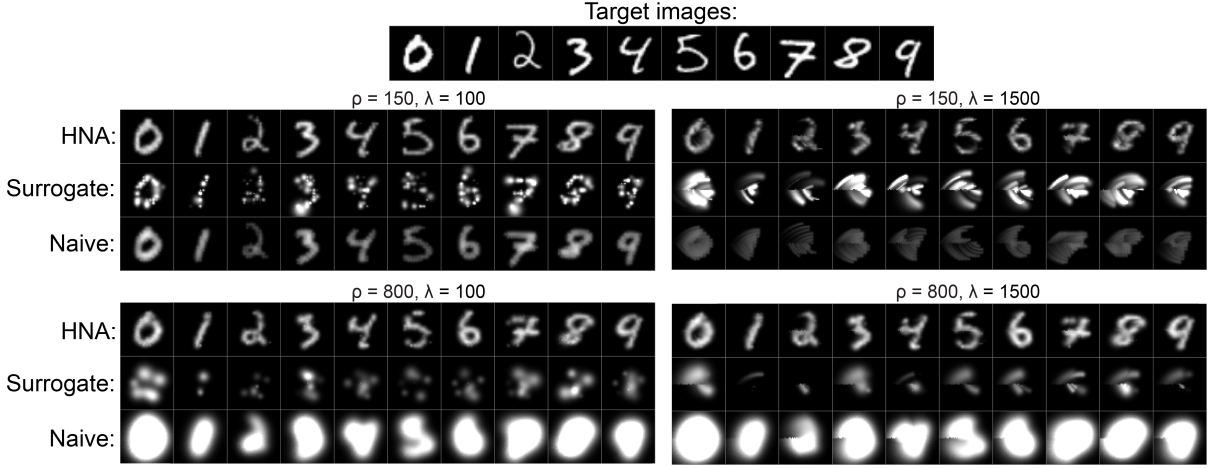


Figure 4.2: Reconstructed MNIST targets for HNA, surrogate, and naive encoders across 4 specific simulated patients. Note that the brightness of the naive encoder is clipped for display

Datasets We first evaluated on handwritten digits from MNIST [deng2012mnist], enabling comparison to previous works [100]. Images preprocessing consisted of resizing the target images to the same shape as the output of f (49x49). We also evaluate on more realistic images of common objects from the MS-COCO [166] dataset. We selected a subset of 25 of the MS-COCO object categories deemed more likely to be encountered by blind individuals (*e.g.* people, household objects), and use only images that contain at least one instance of these objects. We further filter out images by various other criteria, such as being too cluttered or too dim. This process results in a total of approximately 47K training images and 12K test images. See Appendix A.3 for a full description of the selection process.

Natural images often contain too much detail to be encoded with prosthetic vision. While scene simplification strategies exist [90], we focus on the encoding algorithm, so we simply used the ground-truth segmentation masks to segment out the objects of interest. The images were then converted to grayscale, and resized to 49×49 pixels.

Table 4.1: MNIST performance

Encoding	$\rho=150 \ \lambda=100$			$\rho=150 \ \lambda=1500$			$\rho=800 \ \lambda=100$			$\rho=800 \ \lambda=1500$		
	Joint Loss	MAE	RA	Joint Loss	MAE	RA	Joint Loss	MAE	RA	Joint Loss	MAE	RA
Naive	1.161	0.1855	90.3	1.442	0.214	78.1	8.152	1.500	34.8	8.780	1.726	28.8
Surrogate	2.509	0.1351	53.8	3.118	0.2431	30.7	1.692	0.2135	19.9	1.694	0.2237	18.1
HNA	0.559	0.064	98.1	1.029	0.1412	89.3	0.913	0.113	95.9	0.957	0.126	94.8

4.6 Results

4.6.1 MNIST

The phosphenes produced from the HNA, surrogate, and naive encoders on the MNIST test set are shown in Fig. 4.2 and performance is summarized in Table 4.1. For each MNIST sample, the target image is input to the encoder, which predicts a stimulus. The stimulus is fed through the true forward model f , and the predicted phosphene is shown. Since the surrogate method must be retrained for each ϕ , results are only shown for 4 simulated patients. Our proposed approach outperformed the baselines across all metrics (see Appendix A.4 for a comparison of stimuli).

4.6.2 COCO

The phosphenes produced by HNA and the naive encoder for the segmented COCO dataset are shown in Fig. 4.3. We omit the surrogate results due to its poor perceptual performance on MNIST. Averaged across all ϕ , HNA had a joint loss of 0.713 on the test set and MAE of 0.1408, while the naive encoder had a joint loss of 1.873 and MAE of 0.2830.

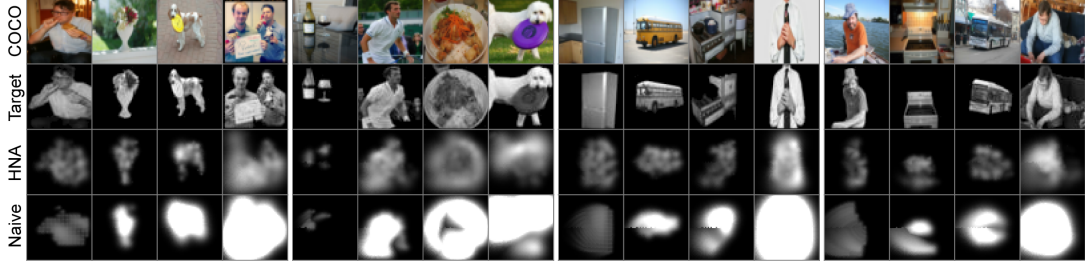


Figure 4.3: Original (*top row*), segmented (*second row*), and reconstructed targets for the COCO dataset, for both HNA (*third row*) and naive encoders (*bottom row*). Left to right within each block of 4 images, ρ takes values of 200, 400, 600, 800. Left to right across blocks, λ takes values of 250, 750, 1250, 2000. Note that the brightness of the naive method is clipped for display.

4.6.3 Modeling Patient-to-Patient Variations

MNIST encoder performance across simulated patients (ϕ) is shown in Fig. 4.4. Since the surrogate encoder has to be retrained for each patient, comparison is infeasible. To visualize the effects of changing ρ and λ on the produced phosphenes, Fig. 4.4A shows the result of encoding two example MNIST digits, both using the naive method and our encoder. As λ increases, the naive phosphenes appear increasingly elongated, and as ρ increases, the phosphenes become increasingly large and blurry. The phosphenes from HNA are slightly too dim and disconnected at low ρ , but are relatively stable across other values of ρ and λ .

To compare performance across the entire dataset, we computed the average test set loss across the same range of ρ and λ (Fig. 4.4B). The encoder performs well across a wide range of simulated patients, with larger loss only at low ρ . The naive method performs well only on a limited set of ϕ , with small λ and $\rho \approx 200$. The naive loss was higher than the learned encoder at every simulated point. Random sampling of ρ and λ for each image results in a joint loss of 0.921, MAE of 0.120, and RA of 94.0% for HNA, while the naive encoder results in a joint loss of 3.17, MAE of 0.596, and RA of 63.6%. The same analysis yielded similar results on COCO (Appendix A.5). An analysis across other

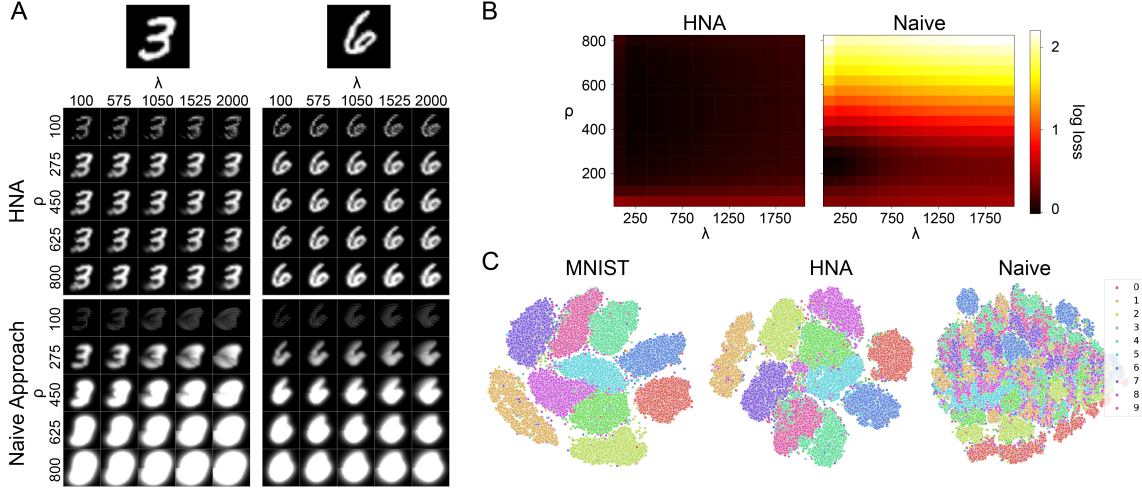


Figure 4.4: Encoder performance across simulated patients (varying ρ and λ) on the MNIST dataset. **A**: Target, HNA encoder, and naive encoder phosphenes for two example digits. **B**: Heatmaps showing the log joint loss across ρ and λ for HNA and naive encoders. **C**: T-SNE clusterings on original MNIST targets, HNA reconstructed phosphenes, and naive reconstructed phosphenes.

parameters is presented in Appendix A.6.

In order for prosthetic vision to be useful, different instances of the same objects would ideally produce similar phosphenes, allowing for consistent perception. To evaluate whether our model achieves this, we cluster the target images and resulting phosphenes using t-SNE [167] shown in Fig. 4.4C. The ground truth images form clusters corresponding to the digits 0-9. The phosphenes from our encoder roughly form similar, slightly less separated groupings, whereas the naive phosphenes do not. To ensure that this was not the result of bad t-SNE hyperparameters, we repeated the clustering across different perplexities and learning rates, obtaining similar or worse results.

4.6.4 Joint Perceptual Error Ablation Study

To show that the joint perceptual metric performs better than any of its individual components, we train models using just the VGG loss and just MAE loss. Shown are values for $\rho=150$ and $\lambda=600$. As mentioned previously, encoders trained using just VGG

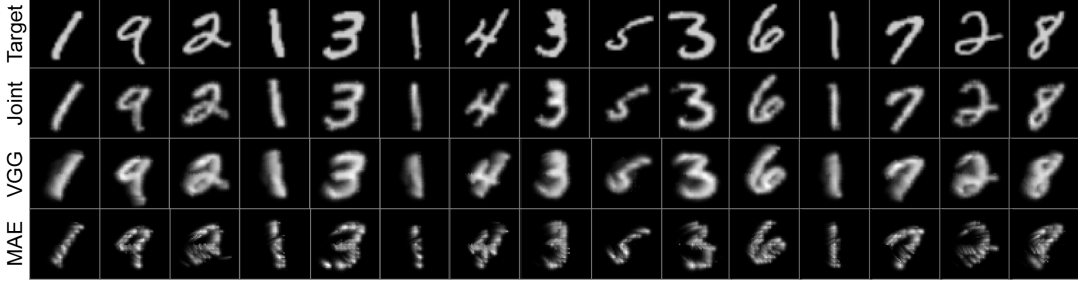


Figure 4.5: MNIST images for HNA encoders trained using the joint, VGG-only, and MAE-only loss.

loss fail to converge, thus we pretrain the VGG encoder using MAE and smoothing loss, then transition to using only VGG. We do not consider ablating the smoothing term (Eq. 4.6) because it is simply a regularization term. Fig. 4.5 shows the phosphenes produced by HNA trained on the joint, VGG-only, and MAE-only loss.

The VGG encoder had a test VGG loss of 4% lower than the joint model, but its produced phosphenes are oversmoothed and blurry. The MAE encoder had a final test MAE of 9% lower than the joint model, but its produced phosphenes are disconnected and low-quality. The joint model had a RA of 99.0%, the VGG encoder had a RA of 95.9%, and the joint model had a RA of 77.6%

4.7 Discussion

Visual Prostheses We found that HNA is able to produce high-fidelity stimuli from the MNIST and COCO datasets that outperform conventional encoding strategies across all tested conditions. Importantly, HNA produces phosphenes that are consistent across representations of the same object (Fig. 4.4C), which is critical to allowing prosthesis users to learn to associate certain visual patterns with specific objects. On the MNIST task, HNA produced high quality reconstructions, nearly matching the targets (Figure 4.2). On the harder COCO task, HNA significantly outperformed the naive encoder, but was

still unable to capture all of the detail in the images. In Appendix A.7, we demonstrate that this is largely due to the implant’s limited spatial resolution and not a fundamental limitation of HNA.

Another advantage of the HNA is that it can be trained to predict stimuli across a wide range of patient-specific parameter values ϕ , whereas the conventional naive encoder works well only for small values of ρ and λ . This may be one reason why the naive encoding strategy has been shown to lead to substantial individual differences in visual outcomes [143, 168]. Our results suggest that stimuli produced with HNA may be able to reduce at least some amount of this patient-to-patient variability.

Furthermore, HNA also proved superior to a surrogate forward model. The latter offer an alternative when the forward model is computationally expensive or not differentiable. Understandably, any inaccuracies in the surrogate model will propagate to the learned encoder during training. However, we observed that even for well trained surrogates, the encoder may still learn to exploit the inexact surrogate instead of learning to invert the true model (see Appendix A.1). It is possible that this exploitation could be mitigated to some extent by adversarially-robust training techniques [169]. We suspect that the surrogate method’s inferior performance here compared to [100] can be explained by our larger stimulus search space. Thus, we cannot currently suggest HNA for surrogate forward models, unless the forward model is sufficiently simple or has a small stimulus space.

Deployment HNA encoders must be lightweight enough to be deployed in resource-limited neuroprosthetic environments. Our encoder’s single image inference time was 1.2ms on GPU and 4ms on CPU. Future work could reduce these numbers through network pruning, mixed precision, and architecture search. Low-power Edge AI accelerators (*e.g.*, Intel’s Neural Compute Stick) and dedicated neuromorphic hardware (*e.g.*, BrainChip’s

Akida SoC) may provide another solution.

Broader Impacts While our work is presented in the context of visual prostheses, the HNA framework may apply to any sensory neuroprosthesis where stimulus selection can be informed by a numeric or symbolic forward model. For example, HNA could be used in cochlear implants [133] to choose stimuli that result in a desired sound, and in spinal cord implants [140] to find the best way to relay neural signals through a damaged section of the spinal cord. Conveniently, the forward models required by HNA have already been developed for a range of applications [150–160]. However, HNA might not apply to all neural interfaces, such as systems without a clear neural or perceptual target (*e.g.*, deep brain stimulation for the treatment of Parkinson’s [170]) or closed-loop systems [141, 171].

Limitations Despite HNA’s potential, the current implementation has a number of limitations. First, as presented the HNA encoder only applies to static targets. Hence dynamic targets must be split into individual frames and encoded separately. However, one approach might be to encode entire stimulus sequences (instead of frames) that are optimized to reconstruct the dynamic target sequence.

Second, HNA works best if there is an accurate forward model mapping from stimulus space to perception. However, Appendix A.8 shows that HNA may still give benefits over a naive encoding even when patient-specific parameters are unknown or mis-specified. In general, if a prosthesis elicits similar results across patients, then a non-patient-specific model would suffice.

Third, the current work deals only with simulated patients. The use of a DNN for stimulus encoding in real patients may raise safety concerns. Since we cannot examine the process by which stimuli are chosen, it is possible that HNA might produce harmful stimuli that could lead to serious adverse events (*e.g.*, seizures). However, this concern is

mitigated by the fact that most neuroprostheses are equipped with firmware responsible for ensuring stimuli stay within FDA-approved safety limits.

4.8 Conclusion

In summary, this paper proposes a hybrid autoencoder structure as a general framework for stimulus optimization in sensory neuroprostheses and, as a proof of concept, demonstrates its utility on the prominent example of visual neuroprostheses, drastically outperforming conventional encoding strategies. This constitutes an important step towards the long-standing challenge of restoring high-quality vision to people living with incurable blindness and may prove a promising solution for a variety of neuroprosthetic technologies.

4.9 Acknowledgements

This work was supported by the National Institutes of Health (NIH R00 EY-029329 to MB).

Chapter 5

Human-in-the-Loop Optimization for Deep Stimulus Encoding

5.1 Summary

Neuroprostheses show potential in restoring lost sensory function and enhancing human capabilities, but the sensations produced by current devices often seem unnatural or distorted. Exact placement of implants and differences in individual perception lead to significant variations in stimulus response, making personalized stimulus optimization a key challenge. Bayesian optimization could be used to optimize patient-specific stimulation parameters with limited noisy observations, but is not feasible for high-dimensional stimuli. Alternatively, deep learning models can optimize stimulus encoding strategies, but typically assume perfect knowledge of patient-specific variations. Here we propose a

⁰This chapter contains work from the author’s previously accepted publication(s): (1) Jacob Granley, Tristan Fauvel, Matthew Chalk, and Michael Beyeler. “Human-in-the-Loop Optimization for Deep Stimulus Encoding in Visual Prostheses”. en. In: *Advances in Neural Information Processing Systems* 36 (Dec. 2023), pp. 79376–79398 (2) Eirini Schoinas, Adyah Rastogi, Anissa Carter, Jacob Granley, and Michael Beyeler. “Evaluating Deep Human-in-the-Loop Optimization for Retinal Implants Using Sighted Participants”. In: *2025 47th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. arXiv:2502.00177 [cs]. IEEE, Apr. 2025

novel, practically feasible approach that overcomes both of these fundamental limitations. First, a deep encoder network is trained to produce optimal stimuli for any individual patient by inverting a forward model mapping electrical stimuli to visual percepts. Second, a preferential Bayesian optimization strategy utilizes this encoder to optimize patient-specific parameters for a new patient, using a minimal number of pairwise comparisons between candidate stimuli. We demonstrate the viability of this approach on a novel, state-of-the-art visual prosthesis model. We show that our approach quickly learns a personalized stimulus encoder, leads to dramatic improvements in the quality of restored vision, and is robust to noisy patient feedback and misspecifications in the underlying forward model. Overall, our results suggest that combining the strengths of deep learning and Bayesian optimization could significantly improve the perceptual experience of patients fitted with visual prostheses and may prove a viable solution for a range of neuroprosthetic technologies.

5.2 Introduction

Sensory neuroprostheses are devices designed to restore or enhance perception in individuals with sensory deficits. They often interface with the nervous system by electrically stimulating neural tissue in order to provide artificial sensory feedback to the user [14, 132]. For instance, visual prostheses have the potential to restore vision to people living with incurable blindness by bypassing damaged parts of the visual system and directly stimulating the remaining cells in order to evoke visual percepts (phosphenes) [18, 45, 48, 143]. However, patient outcomes with current technologies are limited. Patients require extensive training to learn to interpret the evoked percepts, which are typically described as “fundamentally different” from natural vision [59]. Moreover, phosphene appearance varies widely across patients [17], making personalized stimulus optimization

a major outstanding challenge [172].

Much work has gone into developing computational models that can predict the neuronal or perceptual response to an electrical stimulus [17, 22, 76] (often called forward models). Once the forward model is known, a deep neural network can approximate its inverse, thereby identifying the required stimulus to elicit a desired percept [24, 82, 100]. However, these inverse models typically assume perfect knowledge of the patient-specific mapping from stimulation to perception (which is not practically feasible) and are heavily reliant on the forward model’s accuracy over the entire stimulus space.

Alternatively, Bayesian optimization has been successful in personalizing stimulation strategies for many existing neural interfaces [173, 174]. However, this approach is limited in practice because it requires the stimulus dimension to be small (typically < 30 [175], which is orders of magnitudes smaller than the number of stimulus parameters in current implants), and optimization must be repeated for every stimulus. Moreover, visual prosthesis users can typically only give indirect feedback (e.g., verbal phosphene descriptions), unsuitable for traditional Bayesian optimization.

To address these challenges, we propose a novel framework that integrates deep learning-based stimulus inversion into a preferential Bayesian optimization strategy to learn a patient-specific stimulus encoder (Fig. 5.1). First, a DSE is trained to optimize stimuli assuming perfect knowledge of a set of patient-specific parameters (Fig. 5.1, *left*). Second, we embed the DSE within a HILO strategy based on preferential Bayesian optimization, which iteratively learns the ground-truth patient-specific parameters through a series of ‘duels’, where the patient is repeatedly asked their preference between two candidate stimuli. The resulting DSE can then be deployed as a personalized stimulation strategy.

To this end, we make the following contributions:

- We introduce a forward model for retinal implants that achieves state-of-the-art response

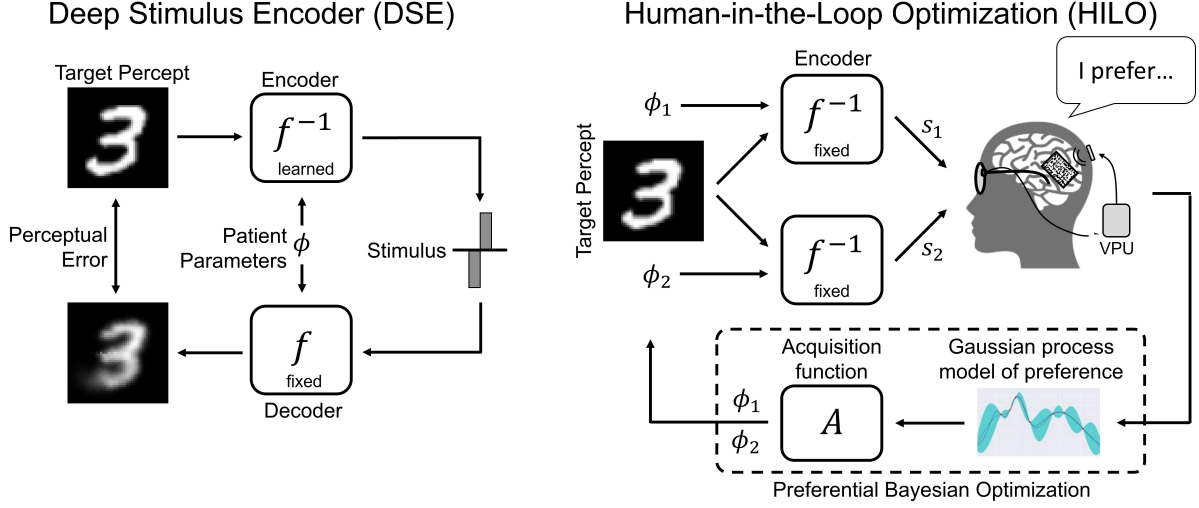


Figure 5.1: *Left*: Deep stimulus encoder (DSE). A forward model (f) is used to approximate the perceptual response to electrical stimuli, subject to patient-specific parameters ϕ . An encoder (f^{-1}) is then learned to minimize the perceptual error between predicted and target percept. *Right*: Human-in-the-loop optimization (HILO). Patient-specific parameters ϕ of the DSE are optimized with user preferences: the patient performs a series of binary comparisons between percepts evoked with different encoders. New pairs of parameters to compare are adaptively selected so as to efficiently find the parameters maximizing the patient’s preference. The target changes each iteration.

predictions. Unlike previous models, this allows us to train a deep stimulus encoder to predict optimal stimuli across 13 dimensions of patient-specific parameters.

- We propose a personalized stimulus optimization strategy for visual prostheses, where a human-in-the-loop optimization (HILO) Bayesian optimization algorithm iteratively learns the optimal patient-specific parameters for a deep stimulus encoder.
- We demonstrate the viability of our approach by conducting a comprehensive series of evaluations on a population of simulated patients. We show HILO quickly learns a personalized stimulus encoder and leads to dramatic improvements in the quality of restored vision, outperforming existing encoding strategies. Importantly, HILO is resilient to noise in patient feedback and performs well even when the forward model is misspecified. Code for the forward model, DSE, and HILO algorithm is available at

<https://github.com/bionicvisionlab/2023-NeurIPS-HILO>.

5.3 Background and Related Work

Visual Neuroprostheses Numerous groups worldwide are pursuing a visual prosthesis that stimulates viable neuronal tissue in the hope of restoring a rudimentary form of vision to people who are blind (Fig. 5.2, *left*) [18, 45, 48, 143]. Analogous to cochlear implants, these devices electrically stimulate surviving cells in the visual pathway to evoke visual percepts (phosphenes). Existing devices generally provide an improved ability to localize high-contrast objects and perform basic mobility tasks.

Much work has focused on characterizing phosphene appearance as a function of stimulus and neuroanatomical parameters [14, 15, 65, 76, 77, 83]. In epiretinal implants, phosphenes often appear distorted due to inadvertent activation of nerve fiber bundles in the optic fiber layer of the retina [17], causing elongated percepts (Fig. 5.2, *center*). In addition, the exact brightness and shape of these elicited percepts depends on the applied stimulus [77] and differs widely across patients (Fig. 5.2, *right*). Granley *et al.* [22]

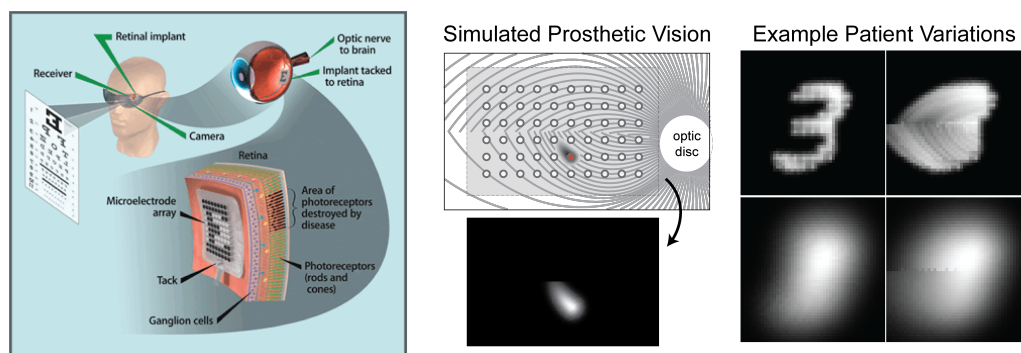


Figure 5.2: *Left*: Visual prosthesis. Incoming target images are transmitted from a camera to an implant in the retina, which encodes the image as an electrical stimulus pattern. *Center*: Electrical stimulation (red disc) of a nerve fiber bundle (gray lines) leads to elongated tissue activation (gray shaded region) and a phosphene (bottom). *Right*: The same stimulus parameters may lead to widely varying visual perceptions in different patients. Adapted with permission from [24].

captured these individual differences by including a set of patient-specific parameters in their phosphene model, denoted by ϕ , which includes both neuroanatomical (e.g., implant location) and stimulus-related parameters (e.g., how brightness scales with amplitude).

Deep Stimulus Encoding Many works attempt to mitigate distortions in prosthetic vision, but do not describe comprehensive stimulation strategies [94, 176, 177]. Those that describe strategies in detail typically require simplification [96] or strong assumptions [178] to be used in practice. Due to the complexities of optimization, deep learning-based stimulus encoders have risen in popularity [24, 82, 100]. In [82], authors proposed an innovative approach where the latent representations of an autoencoder are treated as stimuli and decoded with a phosphene model. However, they used an unrealistic binary phosphene model. Their approach has since been adapted for cortical models [179], and for non-differentiable forward models [100]. Granley *et al.* [24] generalized the approach, showing it could work with realistic forward models across a small range of patients without needing to retrain.

Given a forward (phosphene) model f (mapping stimuli to percepts given ϕ), it is straightforward to show that the optimal stimulus encoder (mapping target images to stimuli) is the pseudoinverse of f [24]. However, to account for the wide range of individual differences in phosphene perception, most realistic forward models are highly nonlinear and not analytically invertible. Thus, previous works have proposed to use the forward model [24, 82] as a fixed decoder within a deep autoencoder trained to minimize the reconstruction error between target images and the predicted percepts. After training, the encoder can be extracted and used to encode target visual inputs in real time. Deep stimulus encoders trained using this approach produce high quality stimuli, but assume knowledge of ϕ . Additionally, if the forward model f is not extremely accurate over the whole stimulus space, then the encoder network might learn to exploit inaccuracies in

the model, producing stimuli that don't generalize to real patients [24]. We utilize an enhanced variant of this approach in our experiments.

Preferential Bayesian Optimization Preferential Bayesian optimization (PBO) is an efficient method for optimizing expensive black-box functions based on binary comparisons [180, 181]. Since the subject's response to stimulation cannot be directly observed, PBO instead builds a Bayesian model of the subject's preferences, g , typically modeled using a Gaussian process. An approximate inference algorithm (expectation propagation; [182, 183]) is used to infer the posterior distribution of the preference function given binary comparison data, $p(g|D)$, which is then used to select new configurations for the next trial according to an acquisition rule. The acquisition rule must balance the exploration-exploitation trade-off inherent to any black-box optimization problem [184].

PBO was previously used to tune BCI stimulation parameters for transcranial [185] and spinal cord stimulation [186]. However, these works directly optimized only a handful of stimulation parameters and cannot translate to visual prostheses, where complex and varying visual inputs have to be mapped to high-dimensional stimuli. To this end, Fauvel & Chalk [187] reduced optimization complexity by inverting a model of phosphene perception, then used PBO to generate encodings preferred by sighted subjects viewing simulated prosthetic vision. However, a linear approximation was used to invert the perception model, which is unrealistic for real-world applications.

Summary We identify 3 main limitations of previous work that this study aims to address:

- **Generalizability of deep stimulus encoders.** Autoencoder-like deep stimulus encoders can accurately optimize stimuli, but require perfect knowledge of patient-specific parameters [24], which can be difficult or impossible to determine in practice [17,

22]. Further, these approaches heavily rely on the accuracy of the forward model [24, 100], while real patients will likely deviate from the forward model. We overcome this limitation by optimizing the learned stimulus encoder based on patients’ preferences, which we show is not bounded by a misspecified forward model.

- **Applicability of Bayesian optimization.** Bayesian optimization is ideally suited for optimizing stimulation parameters based on limited, noisy measurements, but can only optimize a small number of parameters [175]. We train a deep stimulus encoder to output optimal stimuli for any specific patient with known parameters, reducing the search space from the entire stimulus space (hundreds or thousands of parameters) to the low-dimensional space of patient-specific parameters (13 parameters with our model), enabling Bayesian optimization.
- **Simplistic models of perception.** Most previous approaches use overly simplified forward models that do not match empirical data [17, 77]. More accurate models [22] are too computationally expensive to support deep stimulus optimization over a wide range of patients. We overcome this by developing a new phosphene model that matches patient data better than existing models, with significantly reduced memory and time complexity

5.4 Methods

General Framework We consider a system attempting to optimize stimuli for a new patient, specified by a set of (unknown) parameters ϕ . The goal of optimization is a patient-specific stimulus encoder mapping target perceptual responses $\mathbf{t}$ (e.g., visual percepts) to stimuli $\mathbf{s}$.

We assume there exists a forward model f which predicts the patient’s perceptual response to stimulation: $\hat{\mathbf{t}} = f(\mathbf{s}; \phi)$. It follows that the optimal stimulus encoder is the

(pseudo)inverse of f . The inverse can be approximated using an autoencoder-like deep neural network, with f as the fixed decoder and $\hat{f}^{-1}$ as the learned stimulus encoder [24]. The encoder’s weights w are updated using gradient descent to minimize the reconstruction error, measured by some distance function d , between the target $\mathbf{t}$ and the predicted response $\hat{\mathbf{t}}$ averaged across patients and a dataset of targets (Eq. 5.1):

$$w \approx \arg \min_w \mathbb{E}_{\mathbf{t}, \phi} [d(\mathbf{t}, \hat{\mathbf{t}})] \quad (5.1)$$

$$\hat{\mathbf{t}} = f(\hat{f}_w^{-1}(\mathbf{t}, \phi); \phi) \quad (5.2)$$

Once trained, the encoder can accurately predict stimuli, but requires knowledge of the patient-specific parameters ϕ . For a new patient, Bayesian optimization is used to optimize ϕ based on user feedback, thereby learning a personalized DSE. Since the patient’s response cannot be directly measured for visual prostheses, the user is presented with a ‘duel’, i.e. a binary comparison, where they are asked to decide which of two candidate stimuli they prefer [187]. A Gaussian process model of preferences is updated based on the patient’s response, and an acquisition function is used to generate new candidate stimuli. The process can be repeated to iteratively tune the DSE to the patient’s preferences.

Phosphene Model The phosphene model is a differentiable approximation of the underlying biological system (also called a forward model [24]), which maps an electrical stimulus to a visual percept. Although phosphene models exist for visual prostheses, current models either do not match patient data well [17, 76, 77], or are too computationally expensive [22].

Thus, we developed a new phosphene model for epiretinal prostheses (See Ap-

pendix B.1.1 for full details). The model takes in a stimulus vector $\mathbf{s} \in \mathbb{R}^{n_e \times 3}$ specifying the frequency, amplitude, and pulse duration of a biphasic pulse train on each of n_e electrodes. The output phosphene for each electrode is a Gaussian blob centered over the electrode's location $\mu_e(\phi)$ with covariance matrix $\Sigma_e(\mathbf{s}, \phi)$ constructed so that the resulting percept will have area $\rho_e(\mathbf{s}, \phi)$, eccentricity $\lambda_e(\mathbf{s}, \phi)$ and orientation $\theta_e(\phi)$. These functions allow phosphene properties to vary locally with stimulus (e.g., current spread) and anatomical parameters (e.g., electrode location, underlying axon nerve fiber bundle trajectory). The percept for an electrode located at μ_e , is a multivariate Gaussian blob, renormalized to have maximum brightness $b_e(\mathbf{s}, \phi)$:

$$b(x, y) = 2\pi b_e \det(\Sigma_e) \mathcal{N}([x, y]^\top | \mu_e, \Sigma_e), \quad (5.3)$$

where b_e , μ_e , and Σ_e are implicitly parametrized by $\mathbf{s}$ and ϕ . The covariance matrix $\Sigma_e = \mathbf{R}\Sigma_0\mathbf{R}^T$ is calculated from the eigenvalue matrix Σ_0 and a rotation matrix $\mathbf{R}$:

$$\Sigma_0 = \begin{bmatrix} \alpha_e^2 & 0 \\ 0 & \beta_e^2 \end{bmatrix}, \quad R = \begin{bmatrix} \cos \theta_e & -\sin \theta_e \\ \sin \theta_e & \cos \theta_e \end{bmatrix}.$$

The eigenvalues α_e and β_e depend on the intended phosphene area (ρ_e) and elongation (λ_e):

$$\alpha_e^2 = -\frac{\rho_e \sqrt{1 - \lambda_e^2}}{2\pi}, \quad \beta_e^2 = -\frac{\rho_e}{2\pi \sqrt{1 - \lambda_e^2}}.$$

Blobs from individual electrodes are summed into a global percept [63]. Although the sum across electrodes is linear, modulating the size and eccentricity of phosphenes with stimulus parameters makes the final result a nonlinear function of stimulus parameters, preventing analytic inversion. Motivated by previous studies, we used a square 15×15 array of $150\mu\text{m}$ electrodes, spaced $400\mu\text{m}$ apart [24]. In total, the model is parameterized

Table 5.1: Patient-Specific Parameters ϕ

	ρ (dva)	λ	ω	Phosphene Model Parameters						Implant Parameters			
				a_0	a_1	a_2	a_3	a_4	OD_x (μm)	OD_y (μm)	x (μm)	y (μm)	rot (deg)
Lower	1.5	.45	.9	.27	.42	.005	.2	-0.5	3700	0	-500	-500	-30
Upper	8	.98	1.1	.57	.62	.025	.7	-0.1	4700	1000	500	500	30

by 13 patient specific parameters, shown in Table 5.1. The ranges for each parameter were chosen to conservatively encompass all observed patients, centered on the mean value across patients [17, 22, 76, 77, 127].

Deep Stimulus Inversion A deep stimulus encoder (DSE) is a deep neural network responsible for approximately inverting the forward model to produce the optimized stimulus for a target image and a specific patient ($\mathbf{s}_\phi = \hat{f}^{-1}(\mathbf{t}, \phi)$) [24]. We used a network (45M parameters) consisting of blocks, each containing 3 fully connected layers, batch normalization, and a residual connection. The architecture is illustrated in Appendix B.2.1. The flattened target image and the patient specific parameters were passed separately through one block each, concatenated, and passed through another block, after which the amplitude is predicted. The amplitudes were concatenated to the prior intermediate representation, fed through a final block, after which frequency and pulse duration were predicted. The output layers use ReLU activation; all others use leaky ReLU. During training, ϕ were randomly sampled from the range of allowed parameters (Table 5.1). Tensorflow 2.12, an NVIDIA RTX 3090, Adam optimizer, and batch size of 256 [161, 188] were used to train the network.

Human-in-the-Loop Optimization We propose using preferential Bayesian optimization (PBO) to optimize the patient-specific parameters ϕ of the pretrained DSE. Given two sets of patient-specific parameters, ϕ_1 and ϕ_2 , we assume that the probability of a subject preferring ϕ_1 to ϕ_2 (returning a response $\phi_1 \succ \phi_2$) depends on a preference

function $g(\phi)$, modeled using a Gaussian process model:

$$P(\phi_1 \succ \phi_2 | g) = \Phi(g(\phi_1) - g(\phi_2)), \quad (5.4)$$

where Φ is the normal cumulative distribution [189, 190]. The larger the value of $g(\phi_1)$ relative to $g(\phi_2)$, the higher the likelihood that the subject reports preferring ϕ_1 over ϕ_2 .

We used the Maximally Uncertain Challenge [191] to select new comparisons to query, although other popular acquisitions performed similarly (Appendix B.3.3). Searching within the bounds in Table 5.1, this acquisition function selects a ‘champion’, ϕ_1 , which maximizes the expectation of g , and a ‘challenger’, ϕ_2 , for which subjects’ preferences are most uncertain:

$$\phi_1 \rightarrow \arg \max_{\phi} \mathbb{E}_{p(g|D)}[g(\phi)], \quad (5.5)$$

$$\phi_2 \rightarrow \arg \max_{\phi} \mathbb{V}_{p(g|D)}[\Phi(g(\phi) - g(\phi_1))], \quad (5.6)$$

where $\mathbb{V}$ denotes the variance. This algorithm is designed to balance exploitation (values of ϕ that maximize g) and exploration (values of ϕ for which the response is uncertain).

The performance of PBO crucially depends on the Gaussian process kernel and its hyperparameters, which encode our prior assumptions about the latent preference function. Inferring the kernel’s hyperparameters online would slow down the algorithm and could lead to overfitting. Thus, we adopted a transfer learning strategy, which could also be applied to real-life patients (see Appendices B.3.1 and B.3.2 for full details and discussion of alternatives). In brief, for each of 10 patients (with parameters different from those used in the following PBO experiment), we simulated 600 random duels and fit candidate hyperparameters for each of 4 commonly used kernels. We then selected the kernel and hyperparameters that generalized best to the other 9 patients (measured using Brier score

[192] on a held-out test set). The 5/2 Matérn kernel performed best, and was used for all subsequent experiments.

Simulated Patients *In silico* experiments on simulated patients were used to demonstrate the viability of our approach. Each patient was assigned a set of patient-specific parameters ϕ , uniformly sampled from the ranges specified in Table 5.1. When challenged with a duel between two candidate stimuli $\mathbf{s}_{\phi_1}$ and $\mathbf{s}_{\phi_2}$, the simulated patient ran each stimulus through the phosphene model (using ground-truth patient-specific parameters), obtaining the predicted percepts $\hat{\mathbf{t}}_{\phi_1} = f(\mathbf{s}_{\phi_1}; \phi)$ and $\hat{\mathbf{t}}_{\phi_2} = f(\mathbf{s}_{\phi_2}; \phi)$. The users’ preferences were modeled with a Bernoulli distribution, with probability p modulated by the difference in reconstruction error between each percept and the target image:

$$p = \frac{1}{1 + \exp(-\frac{1}{\sigma}(d(\hat{\mathbf{t}}_{\phi_2}, \mathbf{t}) - d(\hat{\mathbf{t}}_{\phi_1}, \mathbf{t})))} \quad (5.7)$$

Here, σ is a configurable parameter that scales the width of the sigmoid, introducing noise into the response. We set σ to be 0.01, chosen empirically based on a conservative estimate: when the error difference was greater than 0.01 it was obvious which percept was better to human observers.

Data and Metrics We used MNIST images as target visual percepts throughout the experiments. Images were resized to be the same size as the output of f (49×49 pixels), and scaled to have a maximum brightness of 2 (aligned with $\text{range}(f)$). Inspired by [24, 82], we used a perceptual similarity metric designed to capture higher-level differences between images [163]. Let $v_l(\mathbf{t})$ be a function that extracts the downstream representations of target $\mathbf{t}$ input to a VGG19 network pretrained on ImageNet [164]. The perceptual

metric is then given by equation 5.8.

$$d(\mathbf{t}, \hat{\mathbf{t}}) = \frac{1}{|t|} (\|\mathbf{t} - \hat{\mathbf{t}}\|_2^2 + \beta \|v_l(\mathbf{t}) - v_l(\hat{\mathbf{t}})\|_2^2) \quad (5.8)$$

This metric was used by the deep stimulus encoder as a training objective, by the simulated patient to choose a duel winner, and throughout HILO as an evaluation metric. $\beta = 2.5\text{e-}5$ was selected via cross-validation (see [24] App. B). To aid in interpretability, we also report a secondary metric based on how identifiable the predicted percepts were. We first pretrained a separate deep net to 99% test accuracy on MNIST classification. We then measured the accuracy of this classifier on the predicted percepts at every iteration of HILO.

5.5 Results

5.5.1 Phosphene Model

To verify that our phosphene model’s predictions line up with observed results from real prosthesis users, we repeated analyses from previous state-of-the-art models, evaluating how phosphene appearance changes with electrode location [17] and stimulus parameters [22, 76, 77, 127]. We used the same datasets, consisting of thousands of phosphene drawings and brightness and size ratings collected across multiple epiretinal prosthesis [45] patients over several years. To evaluate phosphene appearances with electrode location, we calculated the correlation between predicted and observed phosphenes (R_i) for three shape descriptors: area, eccentricity, and orientation. The final score reported is $1 - \sum_i R_i^2$ [17]. To evaluate how phosphene appearance was modulated by stimulus parameters, we calculated the mean squared error between the size and brightness of predicted percepts and patient ratings as amplitude, frequency, and pulse duration were varied. The reported

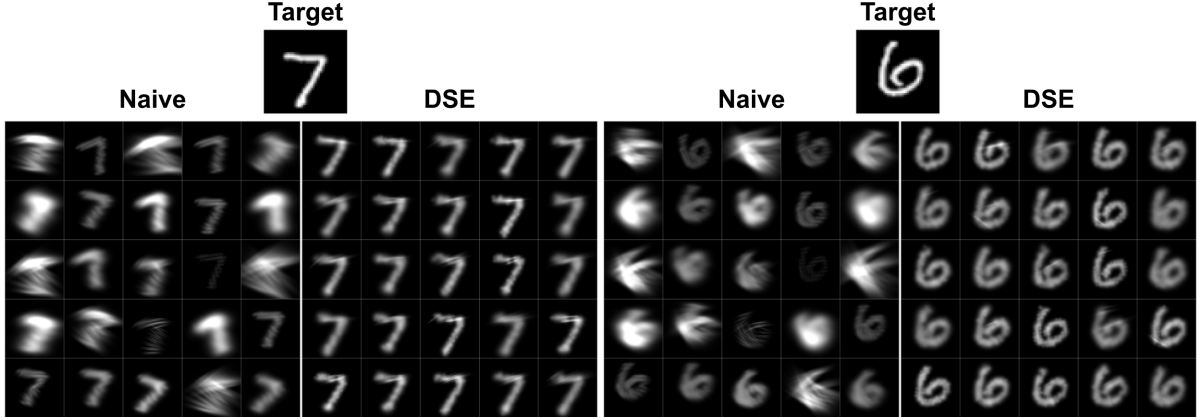


Figure 5.3: Percepts resulting from a naive encoder and the trained DSE for two example target images across 25 randomly selected patients.

values correspond to Figures 4a–c and 5 in [22].

Evaluation results are presented in Table 5.2. Our model significantly outperforms all baselines on the Beyeler *et al.* evaluation, and matches SOTA on the Granley *et al.* evaluations. Additionally, our model is on average 45x faster, and uses 120x less memory than [22]. A detailed description of evaluation methods and additional analysis can be found in Appendix B.1.2.

Table 5.2: Evaluation of Phosphene Model

Model	Electrode Location [17]			Stimulus Parameters [22]			
	S1	S2	S3	4A	4B	4C	5
Nanduri <i>et al.</i> [77]	37.3	19.5	294.8	11.1	72.9	.2	160.9
Beyeler <i>et al.</i> [17]	2.43	7.07	1.15	215.6	108.7	5.52	190.2
Granley <i>et al.</i> [22]	2.43	7.07	1.15	0.9	2.1	0.16	49.5
Proposed	0.28	0.57	0.38	0.73	2.3	0.1	48.6

5.5.2 Deep Stimulus Encoder

We trained a deep stimulus encoder (DSE) to invert our phosphene model (decoder). The encoder was trained across 13 patient-specific parameters, randomly sampled at every

epoch, including for the first time implant position and rotation. This is in contrast to previous DSEs, which either require retraining for every new patient [82, 100, 179], or can only vary two patient-specific parameters [24].

We compared the performance of the DSE to a traditional (‘naive’) encoder [24] currently used by retinal prostheses [45], illustrated in Fig. 5.3. The DSE achieved a test perceptual loss of 0.05 and a MNIST accuracy of 95.6%, significantly outperforming the naive encoder (5.68 and 51% respectively). Note that this performance is when the true patient-specific parameters are known. This performance was slightly better than the values reported in [24] (L1 loss of .108 vs .12 in [24]) despite training across 11 additional patient-specific parameters.

5.5.3 Human-in-the-Loop Optimization

We ran deep learning-based HILO for 100 randomly sampled simulated patients. After every duel, we evaluated the DSE parameterized by the current prediction of patient-specific parameters on a subset of the MNIST test set. The performance of the learned encoder over time (‘HILO’) is illustrated in Figure 5.4, which plots the joint perceptual loss (Figure 5.4.B) and MNIST accuracy (Fig. 5.4C).

As baselines for comparison we used a naive encoder, a non-personalized DSE where the patient-specific parameters are guessed (DSE- ϕ_{Guess}), and an ideal DSE using the true ϕ (DSE- ϕ_{True}). To guess ϕ , we consider two approaches, one which selects the midpoint from the ranges in Table 5.1, and another where performance was averaged across random selections of ϕ from the same ranges. In real patients, we estimate that the performance of a deep stimulus encoder without patient-specific optimization would likely fall somewhere between these two methods, since the distribution of real patients might deviate slightly from Table 5.1. We therefore plot the region bounded by the performance of a DSE with

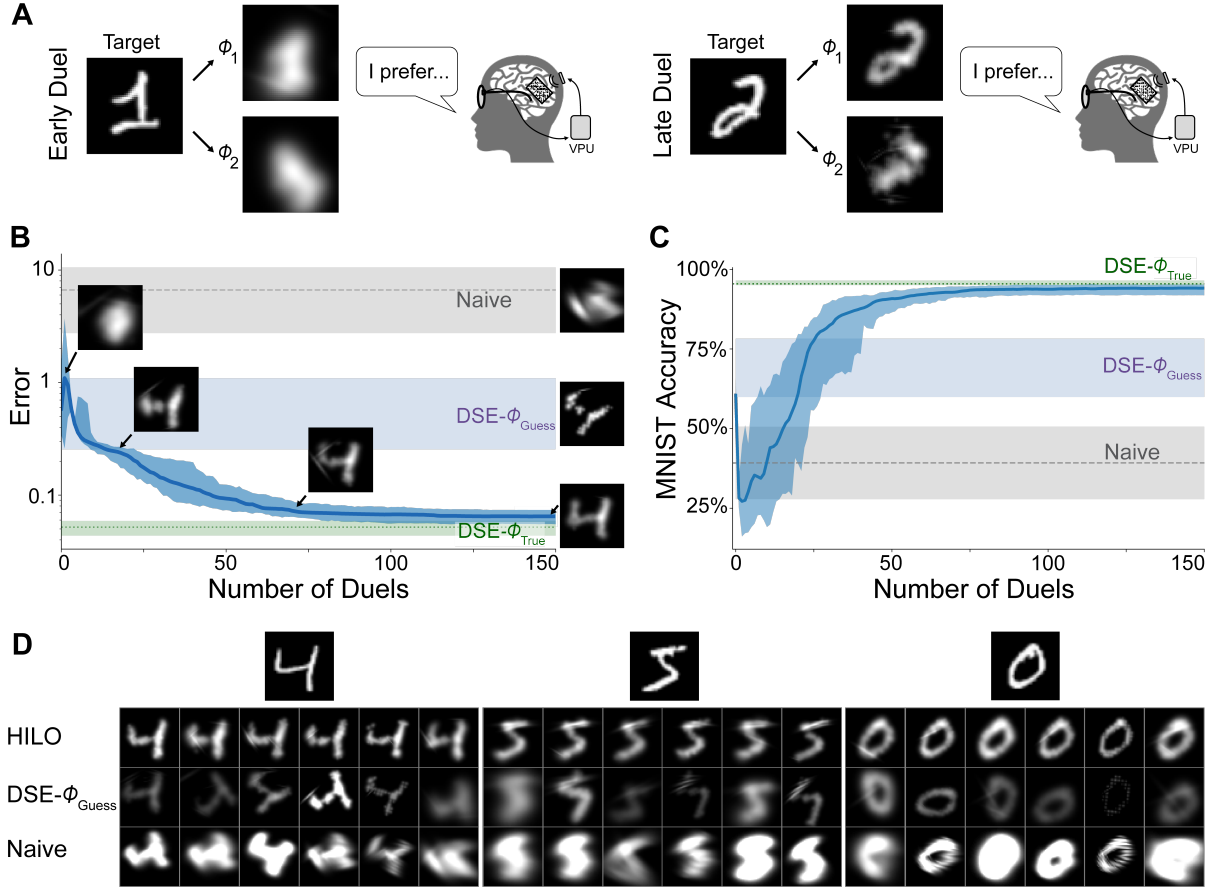


Figure 5.4: Human-in-the-loop optimization of a deep stimulus encoder. **A**: Two example duels, from which patient preferences are learned. **B**: Reconstruction error throughout optimization across 100 simulated patients. Insets show the predicted percept resulting from stimulation with various encoders. Note the y axis is on a log scale. **C**: MNIST accuracy of a pretrained classifier on reconstructed phosphenes. Both plots show smoothed median (window size of 3), with error bars denoting IQR. **D**: Example percepts after optimization for Naive, DSE without HILO, and HILO encoders for 6 random patients. Note, brightness is capped for display for the Naive encoder.

either of these approaches for guessing ϕ . Example percepts after optimization are shown in Fig. 5.4D.

The HILO encoder starts with random predictions, but, after a short initial exploration period, quickly surpassed the baselines. After about 75 iterations, performance approached the ideal DSE encoder, however the HILO encoder still resulted in high-quality percepts after as few as 20 iterations. Averaged across patients, the final reconstruction error

of the HILO encoder was $.071 \pm .0031$ (SEM) and MNIST accuracy was $92\% \pm 1.0\%$. DSE- ϕ_{Guess} had an error of between .25 and 1.1 and MNIST accuracy between 58.6% and 78.3%, and the DSE with true ϕ had an error of $.05 \pm .001$ and accuracy of $95.5\% \pm .1\%$.

5.5.4 Robustness

In reality, it is likely that a patient’s perceptions will not be perfectly captured by the phosphene model. Further, patient responses for visual prostheses are notoriously noisy [59, 193]. To test HILO’s resiliency to these variations, we conducted additional robustness experiments, each with the same 25 simulated patients (Figure 5.5). First, we varied the noise parameter σ in simulated patients’ decision making (Figure 5.5A). Next we constructed various ‘misspecified’ forward models, where the ground-truth model used to decode stimuli differed from the forward model assumed by the DSE. First, we varied the trajectories of the simulated axon bundles [125], thereby changing the orientation of phosphenes (Figure 5.5C). Second, threshold amplitudes for stimulation are notoriously hard to predict, and have been shown to drift by up to 300% over time [194]. Therefore, we tested a variant where the threshold assumed by the encoder was incorrect by up to 300% (Figure 5.5B). Lastly, we used the same forward model, but with patient-specific parameters outside the ranges in 5.1 (Figure 5.5D).

At $\sigma=1e-4$, the patient response was noiseless. For σ equal to .005, .01, and .02, HILO performed similarly to the noiseless model, despite the patient on average making ‘random’ ($p \in [0.35, 0.65]$) decisions in 26%, 38%, and 48% of duels. At $\sigma=0.05$, the decision was ‘random’ 2/3 of the time, and HILO performed similarly or slightly better than the baseline DSE- ϕ_{Guess} . The DSE itself is very resilient to misspecifications in axon trajectory, so HILO performs similarly for this misspecification to the original patients. When thresholds varied, HILO still outperformed the baselines, but converged

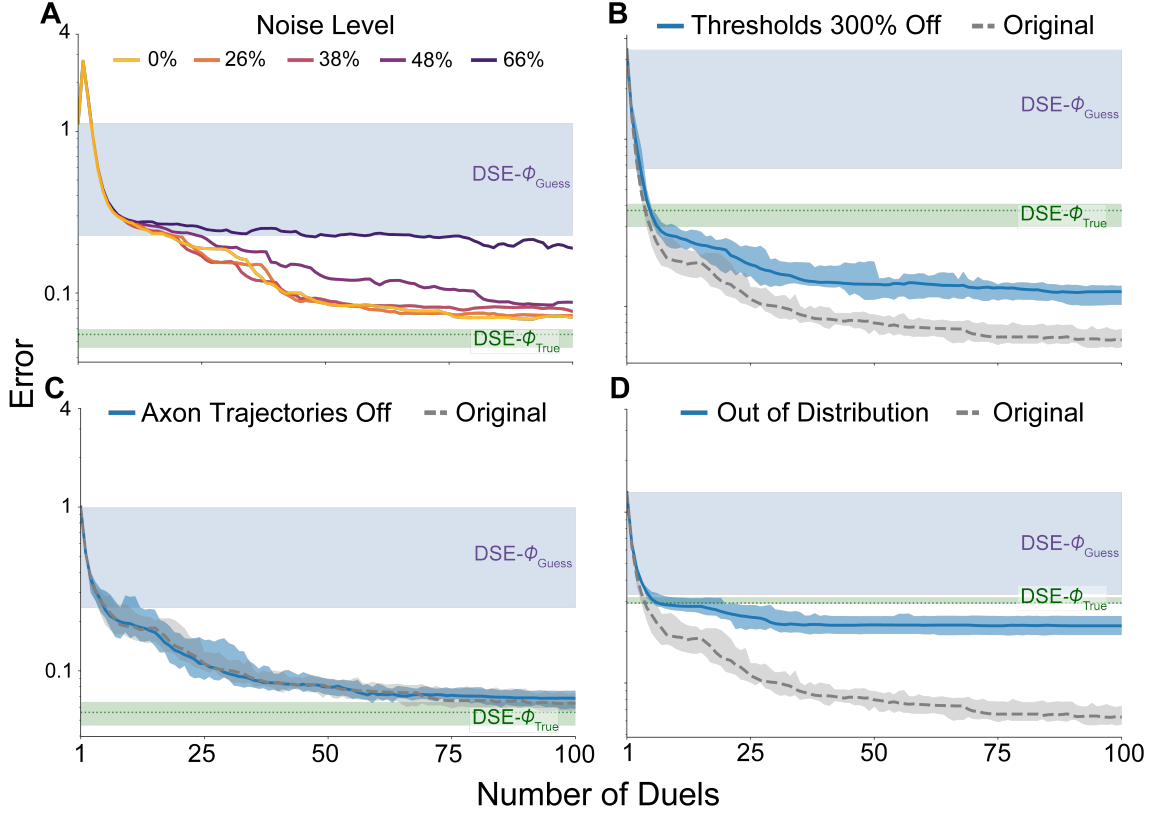


Figure 5.5: Reconstruction error through optimization for noisy patient responses (*upper left*) and for various misspecifications in the forward model assumed by the DSE. Noise level denotes the percentage of duels where the decision was essentially random ($p \in [0.35, 0.65]$), corresponding to σ of $1e-4$, 0.005 , 0.01 , 0.02 , and 0.05 , respectively. All y axes are on log scales. Naive encoders and some error bars omitted for clarity.

to slightly worse encodings than without misspecification. Further, HILO surpassed the DSE encoded with the ground-truth ϕ . This demonstrates that even when the forward model assumed by the DSE is incorrect, HILO still tends to converge to a set of patient-specific parameters that work well for the misspecified patient. For out-of-distribution ϕ , HILO again outperformed both the baseline and true DSEs, but performed worse than in-distribution patients.

5.6 Discussion

Our experiments show that HILO optimization of a deep stimulus encoder led to high-quality, personalized stimulation strategies that outperformed previous state-of-the-art techniques. HILO led to an increase in percept quality compared to using a non-personalized DSE for 99 out of 100 simulated patients, demonstrating the viability of our approach. To enable our HILO algorithm, we also developed a new phosphene model, which is computationally simpler and matches patient data better than previous models, and trained a new DSE, which is able to produce high-quality encodings across all 13 dimensions of patient-specific variations included in our phosphene model. Together, these significantly advance state-of-the-art in patient-specific stimulus encoding, and are important steps towards practically-feasible personalized prosthetic vision in real patients.

The proposed framework combining Bayesian optimization and deep stimulus encoding offers significant improvements over both components in isolation. Use of a DSE allows us to incorporate prior information, reducing the dimensionality of the Bayesian optimization search space from the large stimulus space to the much smaller model parameter space. Our results demonstrate that even when the DSE’s predictions are incorrect, this parameterization is still useful for Bayesian optimization based on patient preferences. Additionally, DSEs are able to invert highly nonlinear forward models, enabling encoder-parameterized Bayesian optimization to be applied to a much larger set of problems. Lastly, the learned encoder can be applied for any target percept, without needing additional optimization. On the other hand, without adaptive feedback from HILO, deep stimulus encoders have no method for learning the individual differences of a new patient, which we show leads to suboptimal stimuli. DSEs rely on the accuracy of their assumed forward model over the entire stimulus space. We show that our approach produces stimuli that work well for the patient, even when the forward model is misspecified, or

when the patient’s responses are noisy.

This approach is practical for stimulus optimization in the wild. The encoder learned during optimization is lightweight, and once deployed, can predict individual stimuli in less than 5 ms on CPU, allowing for high frame rates for prosthetic stimulation. During HILO, updating the Gaussian process model and producing new stimuli on average took 3 seconds, meaning that stimulus optimization could be performed in a matter of minutes. A HILO strategy could be bundled with future visual prostheses, allowing for patients to periodically re-calibrate their devices when they feel the device is not performing adequately, without requiring a clinical professional.

Broader Impacts Although we demonstrate this approach in the context of visual prostheses, our framework is general and could be applied to a variety of sensory devices. Our approach is applicable when the stimulus search space is large and there exists a forward model mapping stimuli to responses. Forward models [150, 156, 157, 160] and deep stimulus encoders [114, 115, 195] have been successfully used across multiple sensory modalities, and could potentially be adapted for personalization with HILO.

Limitations Although promising, our approach is not without limitations. We assumed that the preference of patients for different stimuli is related to the distance metric used to measure perceptual similarity, which may not be true in practice. However, results by [187] suggest that PBO is robust to a mismatch between the distance metric used to invert the forward model and the preference of patients. Another limitation is that evaluation of our approach was only performed on simulated patients with a simulated perceptual model. However, this is mitigated by the fact that HILO showed robustness to model inaccuracies. Still, since it is difficult to predict the behavior of deep learning models, using a deep stimulus encoder in real patients could raise safety concerns. It

may be possible for a deep encoder to produce unconventional stimuli, potentially leading to adverse effects. However, most devices come with firmware responsible for ensuring stimuli stay within FDA-approved safety limits.

In conclusion, our results suggest that combining the strengths of deep learning and Bayesian optimization could significantly improve the perceptual experience of patients fitted with visual prostheses and may prove a viable solution for a range of neuroprosthetic technologies.

5.7 Follow-up in Sighted Humans¹

So far, HILO results have been presented in the context of stochastic simulated patients. However, these simulations relied on predefined probabilistic rules, where choices were more likely when a percept had a lower computed error metric. Human decision making, by contrast, is inherently more variable and may not align with model assumptions, leaving open the question of whether HILO would perform similarly when real users provide feedback.

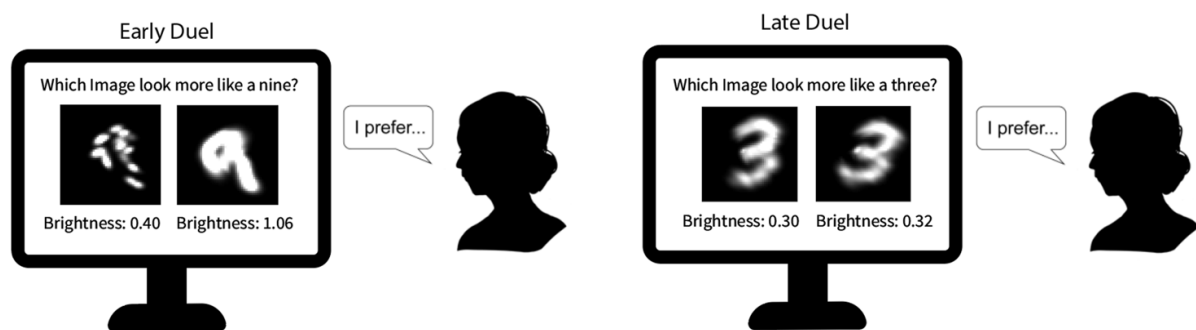


Figure 5.6: Examples of the duels presented to sighted participants on a computer monitor. Participants selected the preferred stimulus based on both perceived shape and reported numerical brightness.

¹This section was based on work from the author’s publication: Eirini Schoinas, Adyah Rastogi, Anissa Carter, Jacob Granley, and Michael Beyeler. “Evaluating Deep Human-in-the-Loop Optimization for Retinal Implants Using Sighted Participants”. In: *2025 47th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. arXiv:2502.00177 [cs]. IEEE, Apr. 2025

Next, we take an intermediate step toward real-world validation by incorporating sighted participants viewing SPV stimuli on a monitor (Fig. 5.6). This approach enables systematic testing of whether HILO can adapt to individual perceptual variability, maintain robustness to model misspecifications, and remain practical under real-world decision-making constraints. Importantly, we evaluated HILO not only under standard conditions but also in out-of-distribution scenarios, testing its generalizability beyond the training range. We also systematically varied key phosphene properties, such as size and elongation, highlighting how perceptual distortions affect optimization performance and motivating future efforts to improve phosphene focality.

5.7.1 Methods

Participants

Seventeen sighted undergraduate students from the University of California, Santa Barbara, participated in the study (10 female, 7 male; ages 18–21; $M = 19.6$, $SD = 1.12$). Participants were recruited through university-wide announcements and provided informed consent prior to participation. All participants reported normal or corrected-to-normal vision and no history of neurological or visual impairments. Participants were briefed on the study’s purpose and tasks before beginning the experiment.

Task

Participants viewed simulated prosthetic vision on a monitor and completed a series of trials aimed at optimizing the mapping between simulated electrical stimuli and perceived visual representations. Each trial, referred to as a “duel”, presented two full percepts generated for a target image. Participants selected the percept they judged to better match the target, which was described textually (e.g., “number eight”) to encourage

independent mental representations. To prevent selection bias, the positions of the two stimuli (left or right) were randomized across trials.

In practice, electrical stimulation can evoke phosphenes that are excessively bright or even painful [9, 43, 44, 118]. To account for the limited dynamic range of computer monitors, numerical brightness values were displayed alongside each stimulus. During the tutorial phase, participants were introduced to the brightness scale: 1 represented threshold brightness (too dim), 2 was twice the threshold (ideal), and higher values indicated increasing excessive brightness (Fig. 5.7).

The experiment consisted of three phases:

- *Tutorial phase:* Participants completed practice trials to familiarize themselves with the task, including how to make selections and interpret brightness values.
- *Optimization phase:* Participants completed 60 duels, during which the HILO framework iteratively refined the simulated user parameters ϕ based on their preferences.
- *Evaluation phase:* Participants completed 39 additional duels, comparing the optimized DSE against a naïve baseline encoder.

5.7.2 Experimental Conditions

The study included three experimental conditions, designed to evaluate the performance and robustness of the HILO framework under varying levels of complexity and realism:

- *Main experiment:* This was the standard experiment where the unknown parameters ϕ for simulated users were iteratively optimized using participant feedback, following the HILO procedure described in Section 5.5.3.

- *Threshold misspecification (TM)*: This experiment introduced errors in the assumed threshold amplitudes of up to 300%, testing the framework’s robustness to inaccuracies in user-specific parameters.
- *Out-of-distribution (OOD)*: In this experiment, the ground-truth ϕ values for simulated users were drawn from a distribution outside the training range of the deep stimulus encoder (DSE), evaluating the framework’s ability to generalize to unseen parameter configurations.

In each condition, the HILO framework was evaluated against two baseline models:

- *Naïve encoder*: The approach traditionally used by prostheses, where the target image is reduced to the electrode array resolution, and each pixel’s grayscale value is directly scaled to stimulus amplitude.
- *Deep stimulus encoder (DSE)*: The same deep stimulus encoder as used in HILO, but instead of user-specific parameters being tuned, they are guessed as the mean of the observed ranges reported in 5.1 (equivalent to $\text{DSE-}\phi_{\text{Guess}}$ in Section 5.5.3).

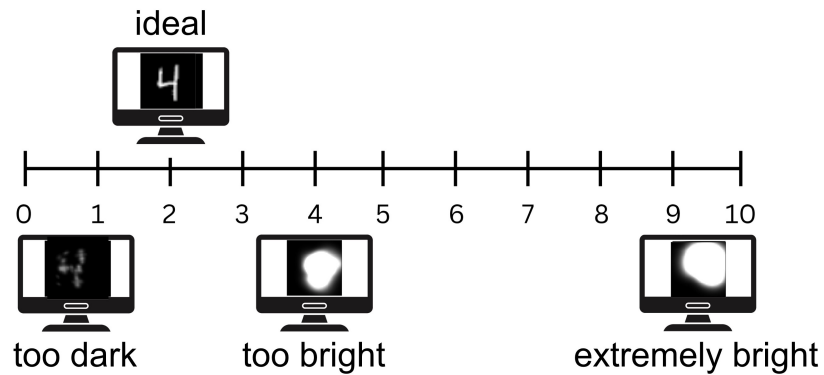


Figure 5.7: Example stimuli illustrating the range of phosphene brightness levels shown to participants. A value of 0 represents complete darkness, 2 is the ideal brightness for a retinal prosthesis user, 5 is overly bright, and 10 is extremely bright, with white filling most of the stimulus area.

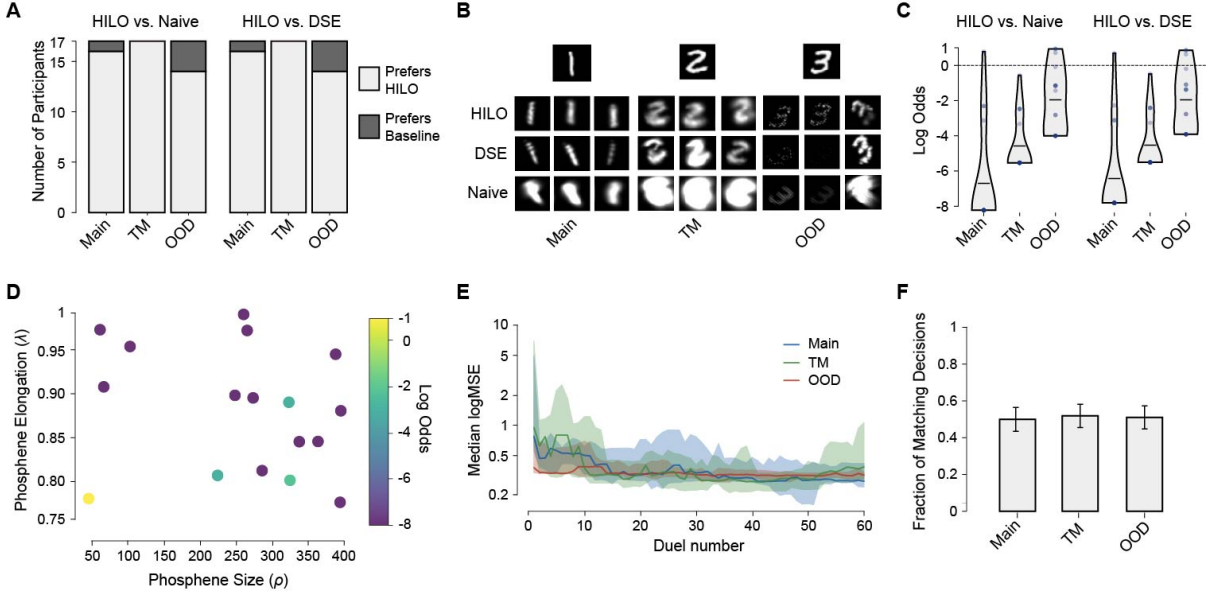


Figure 5.8: **A)** Number of participants who significantly preferred human-in-the-loop optimization (HILO) over the naïve encoder (left) and the deep stimulus encoder (DSE) without HILO (right), based on log odds less than 0 in a linear mixed-effects model. **B)** Example percepts generated by the HILO encoder, the DSE without HILO, and the naïve encoder for three participants across the three experimental conditions. **C)** Distribution of log odds for HILO across the three experiments. **D)** Distribution of the two main user-specific parameters, ρ (phosphene size) and λ (axon-aligned elongation), colored by the log odds indicating preference for HILO in the main experiment. **E)** Median mean squared error (MSE) over the course of optimization for each experiment, with shaded regions denoting the interquartile range (IQR). **F)** Proportion of duels where participant decisions matched those of the simulated agent.

5.7.3 Results

Participant Preferences for HILO vs. Other Encoders

Across all three experimental conditions, participants consistently preferred HILO-optimized stimuli over both the naïve encoder and the non-personalized DSE- ϕ_{Guess} (Fig. 5.8A). Example percepts across the three experimental conditions are shown in Fig. 5.8B. Statistical analysis was performed with a linear mixed-effects logistic regression model, with population and per-subject effects. Across subjects, HILO was significantly preferred over the naïve and DSE encoders for all three experiments ($p < 0.0001$). In

the main experiment, 16 out of 17 participants favored HILO over the other encoders (defined as having a subject-level log odds of less than 0, indicating preference for HILO). In the threshold misspecification condition, all participants selected HILO. In the out-of-distribution condition, 14 out of 17 participants preferred HILO.

Percepts generated by HILO generally retained greater structure and recognizability compared to baseline methods. In the main experiment, both HILO and DSE produced structured percepts that aligned with target images, whereas the naïve encoder often resulted in highly distorted or unrecognizable shapes. Under threshold misspecification, HILO maintained consistent percepts, whereas DSE outputs varied widely, sometimes producing overexposed phosphenes. In the OOD condition, distortions appeared across all methods, though HILO percepts remained more structured and interpretable than those generated by the alternatives.

The distributions of log odds per participant are shown in Fig. 5.8C. The mean log odds for HILO over the naïve encoder were -6.70 ± 0.71 (main experiment), -4.57 ± 0.39 (threshold misspecification), and -1.95 ± 0.46 (OOD condition). Lower log odds indicate stronger preference for HILO, confirming its advantage across conditions. Although more participants chose HILO in the threshold misspecification experiment than in the main experiment, their preference was weaker, as reflected in the less negative log odds.

Log Odds Across Individual User-Specific Parameters

To assess how individual user-specific factors influenced preference trends, log odds were examined as a function of ρ (phosphene size) and λ (phosphene elongation) Fig. 5.8D. Participants with larger ρ and λ , corresponding to larger, and more elongated phosphenes, exhibited higher log odds in favor of HILO, while those with smaller, minimally distorted percepts showed lower log odds differences between encoders. This pattern suggests that HILO provides the greatest benefit when percepts are highly distorted but offers less

improvement when phosphenes are already relatively structured.

Comparison Between Human and Simulated Preferences

MSE decreased over the course of optimization (Fig. 5.8E), confirming that participant choices contributed to improved stimulus encoding. The final loss in the main experiment converged to 0.27, with similar performance in the other two conditions. In contrast, simulated patients achieved a much lower final loss of 0.07 5.4. The slower rate of loss reduction in human trials suggests that participants were not strictly optimizing for pixel-wise reconstruction accuracy, instead incorporating additional perceptual or cognitive factors into their choices.

To further quantify these differences, we compared participant selections to those predicted by the simulated agents. (Fig. 5.8F). Across all three experiments, participants made the same choice as the simulated agent in only about 50% of trials, indicating that human decision-making diverges from the assumptions made in purely simulation-based studies. Despite this variability, participants still overwhelmingly preferred HILO, demonstrating its robustness to real-world human responses and reinforcing the need for empirical validation beyond theoretical models.

5.7.4 Discussion

HILO Stimuli Are Preferred Across Conditions

This follow-up study demonstrates that human-in-the-loop optimization (HILO) successfully personalizes deep stimulus encoders (DSEs) for individual users, even in the presence of model inaccuracies and noisy human feedback. Sighted participants viewing simulated prosthetic vision consistently preferred HILO-optimized stimuli over both naïve encoding and a standard DSE across all experimental conditions, reinforcing the viability

of this approach for adapting neuroprosthetic devices to individual users. These findings provide critical evidence supporting the application of HILO to real-world neuroprosthetic calibration.

Human Decision-Making Differs from Simulated Users

The simulated patients introduced in 5.5.3 attempted to account for human variability by introducing artificial noise (i.e., HILO remained effective even when simulated users made random choices in two out of three trials), however, real human decision-making may not align with these assumptions. The experiments in human subjects confirm that human choices diverged from simulated predictions.

Although MSE decreased over the course of optimization, the final loss was significantly higher than in purely simulated experiments, suggesting that human participants did not strictly optimize for pixel-wise accuracy. Additionally, participants selected the same choices as the simulated agent in only about 50% of trials. This suggests that human decision-making incorporates perceptual factors beyond those captured by the computational model. Despite this divergence, participants still overwhelmingly preferred HILO, reinforcing the need for validation with real users rather than relying solely on theoretical models.

Limitations and Future Work

While this study provides strong support for the use of HILO in personalizing neuroprosthetic devices, several important challenges remain. The next step is testing HILO with blind prosthesis users, though this presents practical and safety considerations. Deep learning models can produce unpredictable outputs, and while most neuroprosthetic devices include firmware safeguards to enforce stimulation limits, additional validation will be needed to ensure that optimized stimuli do not generate unintended or adverse

effects.

Additional limitations of the present study include the use of a static computer screen to display simulated percepts, which does not fully capture the dynamic experience of artificial vision in everyday life. Moreover, the simulated percepts shown to participants did not incorporate the temporal fading that real prosthesis users experience, potentially affecting the ecological validity of participant responses.

Future work should refine the optimization objective to better align with human perceptual judgments, moving beyond pixel-wise error metrics toward loss functions that capture higher-level visual features [196]. Extending validation to blind participants and testing under more dynamic, real-world conditions will be critical for clinical translation. Recent work underscores the need to align prosthetic development with user needs [20] and the realities of assistive technology use in daily life [197], emphasizing that future optimization must prioritize not only perceptual similarity but also usability and functional task performance. Additionally, optimizing stimulation to align with latent neural representations in the visual cortex [198] offers a promising path toward deeper integration between artificial stimulation and biological processing. Together, these directions move beyond technical optimization toward meaningful, user-centered neuroprosthetic design.

Beyond visual prostheses, the HILO framework may generalize to other sensory neuroprosthetic systems, such as cochlear implants and tactile feedback devices, where forward models and deep encoders are increasingly being adopted [115, 150, 157, 199]. Recent work on brain co-processors [200] emphasizes the growing importance of closed-loop systems that jointly optimize neural decoding and stimulation in collaboration with the brain itself. Our findings support this broader vision by demonstrating that real human feedback can successfully guide the adaptation of stimulus encoding strategies, advancing toward tighter integration of artificial and biological neural systems. By establishing the viability of HILO with human participants, this study represents an important step

toward fully individualized optimization of neuroprosthetic technologies.

5.8 Acknowledgements

Research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under Award Number DP2LM014268. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Chapter 6

Deep Learning–Based Control of Electrically Evoked Activity in Human Visual Cortex

6.1 Summary

Visual cortical prostheses offer a promising path to sight restoration, but current systems elicit crude, variable percepts and rely on manual electrode-by-electrode calibration that does not scale. This work introduces an automated data-driven neural control method for a visual neuroprosthesis using a deep learning framework to generate optimal multi-electrode stimulation patterns that evoke targeted neural responses. Using a 96-channel Utah electrode array implanted in the occipital cortex of a blind participant, we trained a deep neural network to predict single-trial evoked responses. The network was used in two complementary control strategies: a learned inverse network for real-time stimulation synthesis and a gradient-based optimizer for precise targeting of desired neural responses. Both approaches significantly outperformed conventional methods in controlling neural

activity, required lower stimulation currents, and adapted stimulation parameters to resting state data, reliably evoking more stable percepts. Crucially, recorded neural responses better predicted perceptual outcomes than stimulation parameters alone, underscoring the value of our neural population control framework. This work demonstrates the feasibility of data-driven neural control in a human implant and offers a foundation for next-generation, model-driven neuroprosthetic systems, capable of enhancing sensory restoration across a range of clinical applications.

6.2 Introduction

Implanted neuroprostheses offer a promising route to restoring lost sensory function by electrically stimulating neurons in brain areas responsible for perception. Devices targeting vision [1–3], audition [4], and somatosensation [5, 6] are in various stages of development, yet all face a fundamental challenge: interfacing with highly non-linear networks of biological neurons whose role in perception is not fully understood. Due to the limited spatial resolution of electrical stimulation, prostheses evoke neural response patterns that are unnatural to the brain, often resulting in percepts that are artificial, distorted, or variable [9–12].

One prominent example of neuroprostheses is the visual prosthesis, which electrically stimulates neurons in the retina [45–48], optic nerve [49], lateral geniculate nucleus [50], or early visual cortex [18, 19, 51, 52] to restore visual perception. Cortical visual prostheses, which bypass the eye and optic nerve entirely, can both stimulate and record neurons in the primary visual cortex using microelectrode arrays such as the Utah Electrode Array (UEA) [18, 54]. This bidirectional interface enables the delivery of patterned stimulation and the measurement of evoked neural responses (Figure 6.1A).

Despite their potential, cortical implants to date have rarely elicited percepts beyond

simple shapes or flashes of light [9, 18, 19, 52]. Phosphenes (i.e., percepts induced by electrical stimulation) are often distorted, difficult to interpret, and inconsistent. Their size, brightness, shape, and color vary across time, electrodes, and individuals [18, 19]. Percepts from simultaneous multi-electrode stimulation typically do not correspond to the linear sum of individual phosphenes [18, 52], reflecting complex cortical connectivity [66, 67] and nonlinear neural integration [201]. Selecting stimulation patterns that produce meaningful and stable percepts is, therefore, nontrivial.

The conventional ‘1-to-1 mapping’ strategy assumes a direct visuotopic correspondence between visual space and the electrode grid, mapping camera pixel intensities directly to stimulation amplitudes. This ignores nonlinearities in cortical responses and perception, often resulting in distorted percepts. While alternative stimulation strategies have been proposed [24, 25, 82, 93–96, 98–100, 102, 112, 202–204], none have been validated in vivo in humans using perceptually relevant pulse-train stimulation, and most do not address the substantial day-to-day variability in cortical responses that limits clinical reliability.

Recent evidence points to a tight coupling between neural responses and perception. In macaques, intracortical electrode recordings in V4 predicted the current threshold for perception [205], and cortical microstimulation could bias visual perception [206]. In humans, recorded neural activity from a UEA correlated with perceived brightness, detection thresholds, and percept duration [73, 207]. Additionally, pre-stimulus activity has been shown to influence both neural responses and perceptual thresholds [73, 208–210]. These findings suggest that closed-loop approaches, which monitor and adaptively control neural activity, could improve the consistency and quality of stimulation-evoked perception [172].

Here, we introduce a deep learning-based framework for neural activity shaping [112] in a human cortical visual prosthesis. Using multi-session recordings from a participant implanted with a UEA in early visual cortex (Figure 6.1A), we trained a forward neural

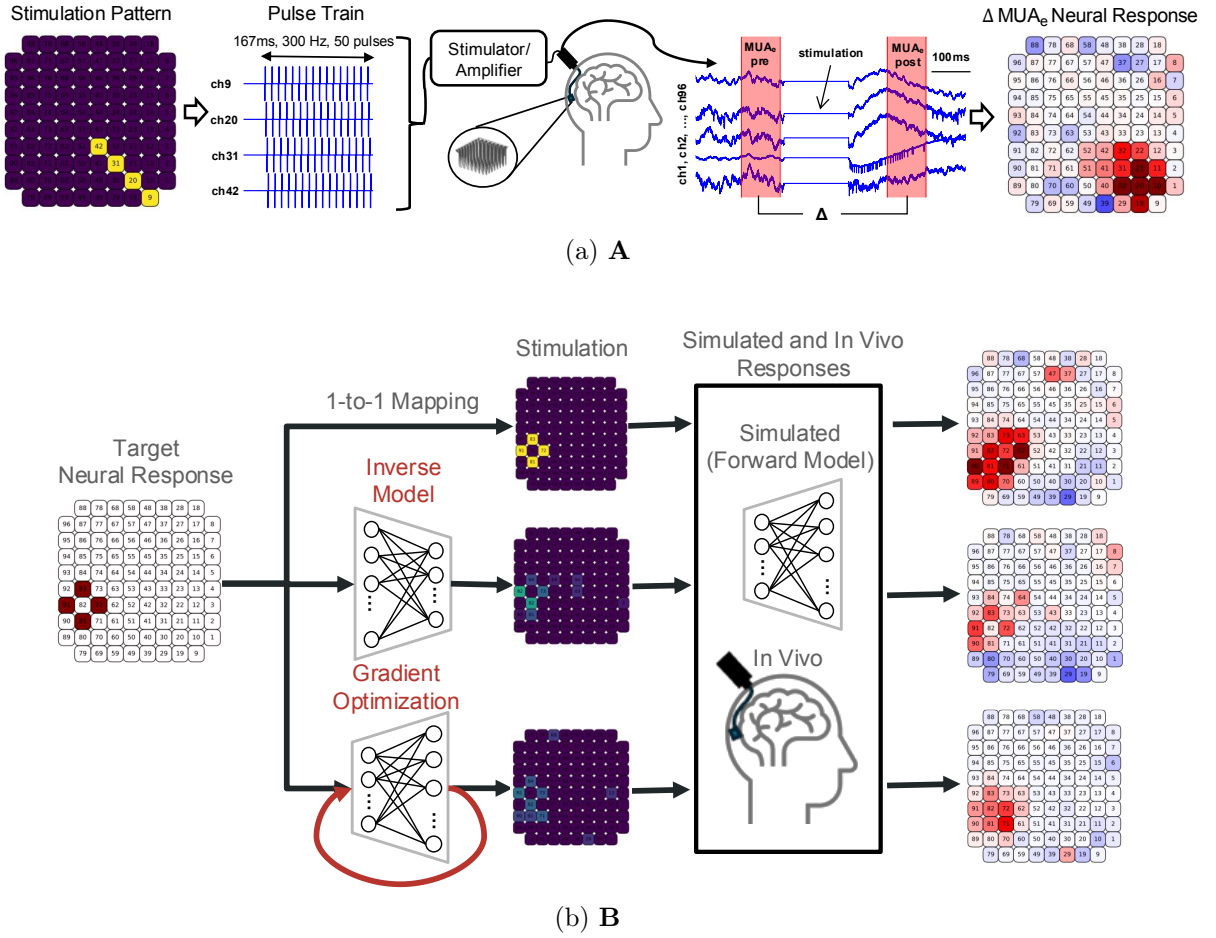


Figure 6.1: **Schematic representation of the experimental setup and modeling framework.** **A)** A stimulation pattern across electrodes is selected and sent to a UEA, which delivers the stimulus to the primary visual cortex of a human participant. The electrical activity under each electrode is recorded, amplified, and multi-unit spiking activity is extracted. **B)** Visualization of three approaches for neural response targeting: the conventional 1-to-1 mapping, the proposed inverse neural network method, and the proposed gradient-based optimization. Generated stimuli can then be played *in vivo* or input to the forward model to obtain real and simulated neural responses.

network (NN) to predict single-trial evoked responses from stimulation parameters, incorporating pre-stimulus activity to account for day-to-day drift. This model served as an *in silico* testbed for two complementary control strategies: (1) a gradient-based optimizer that iteratively refines stimulation parameters to minimize response error, and (2) a learned inverse model that generates stimulation patterns from target neural

responses in real time (Figure 6.1B). In vivo experiments demonstrate that both approaches improve control of population responses compared to 1-to-1 mapping and linear baselines, elicit more consistent percepts, and require lower currents. These findings establish the feasibility of model-driven neural control in a human visual cortical implant, representing a step toward closed-loop, scalable, and perceptually informed cortical visual prostheses.

6.3 Results

6.3.1 Deep Learning Model of Neural Responses to Electrical Stimulation

We first sought to model the relationship between electrical stimulation and evoked neural responses in human visual cortex. To this end, we collected a large dataset of stimulation-evoked activity from a 96-channel Utah Electrode Array (UEA) implanted near the primary visual cortex of a blind participant.

Stimulation patterns spanned both random ($N=5818$, Supplementary Figure 1A) and structured ($N=484$, Supplementary Figure 1B) multi-electrode configurations and were delivered over the course of four months across 26 separate days. On a subset of trials ($N=5515$), the participant was instructed to report whether they perceived a phosphene ($N=4811$) and to describe it (shape, color, brightness, size; $N=704$). Neural responses to stimulation were recorded using the same Utah Electrode Array, from which we extracted the envelope multi-unit activity (MUA_e). This signal reflects the pooled spiking activity of multiple neurons near each electrode tip and provides a robust estimate of local population activity [205]. To reduce variability introduced by baseline fluctuations and highlight stimulation-evoked changes, we computed a summary measure we refer to as change in envelope multi-unit spiking activity (ΔMUA_e): the difference between the mean MUA_e

in a post-stimulation window (100–200 ms after stimulus offset) and a pre-stimulation baseline (-110 to -10 ms). These intervals were selected to avoid contamination by stimulation artifacts while capturing short-latency evoked responses. Because responses to identical stimulation patterns varied substantially across days, we also incorporated average pre-stimulus activity from each day to account for state-dependent drift and improve across-session generalization. For more details see Methods 6.5.1-6.5.4.

The same stimulation generally evoked similar responses but with day-to-day variations (Figure 6.2A). For repeated stimulation patterns, the neural response’s variance across days was significantly larger than the variance within a day (Figure 6.2B, $p < 0.0001$, Wilcoxon signed-rank test; see Methods 6.5.7 for details on statistical tests). We found that ΔMUA_e exhibited decreased variability to repeated stimuli across days than post-stimulation MUA_e (Figure 6.2C, $p < 0.001$), suggesting that it is less dependent on daily short-term variations in neural state (see Supplementary Figure 3A-C). Principal Component Analysis revealed that only 48 components were required to explain 95% of the variability in ΔMUA_e , whereas the stimulation space required all dimensions (Figure 6.2D). This suggests that neural responses are constrained to a lower-dimensional manifold, whereas the stimulation space is high-dimensional.

From the dataset of neural responses, a deep neural network (DNN) (1.2M parameters, ‘Forward NN’, see Methods 6.5.8) was trained to predict the neural response (ΔMUA_e) that would result from a given stimulation pattern. Since neural activity varied significantly day-to-day, the network was also given as input the average pre-stimulation (i.e., resting state) activity for each day. The model was trained using data from 19 days and evaluated on held-out data from 5 random non-consecutive days, resulting in a total of 5123 and 1179 train and test responses, respectively. The DNN was trained using gradient descent to minimize the MSE between the predicted and true responses.

The forward NN was compared against six baseline models (Methods 6.5.8): (1) a

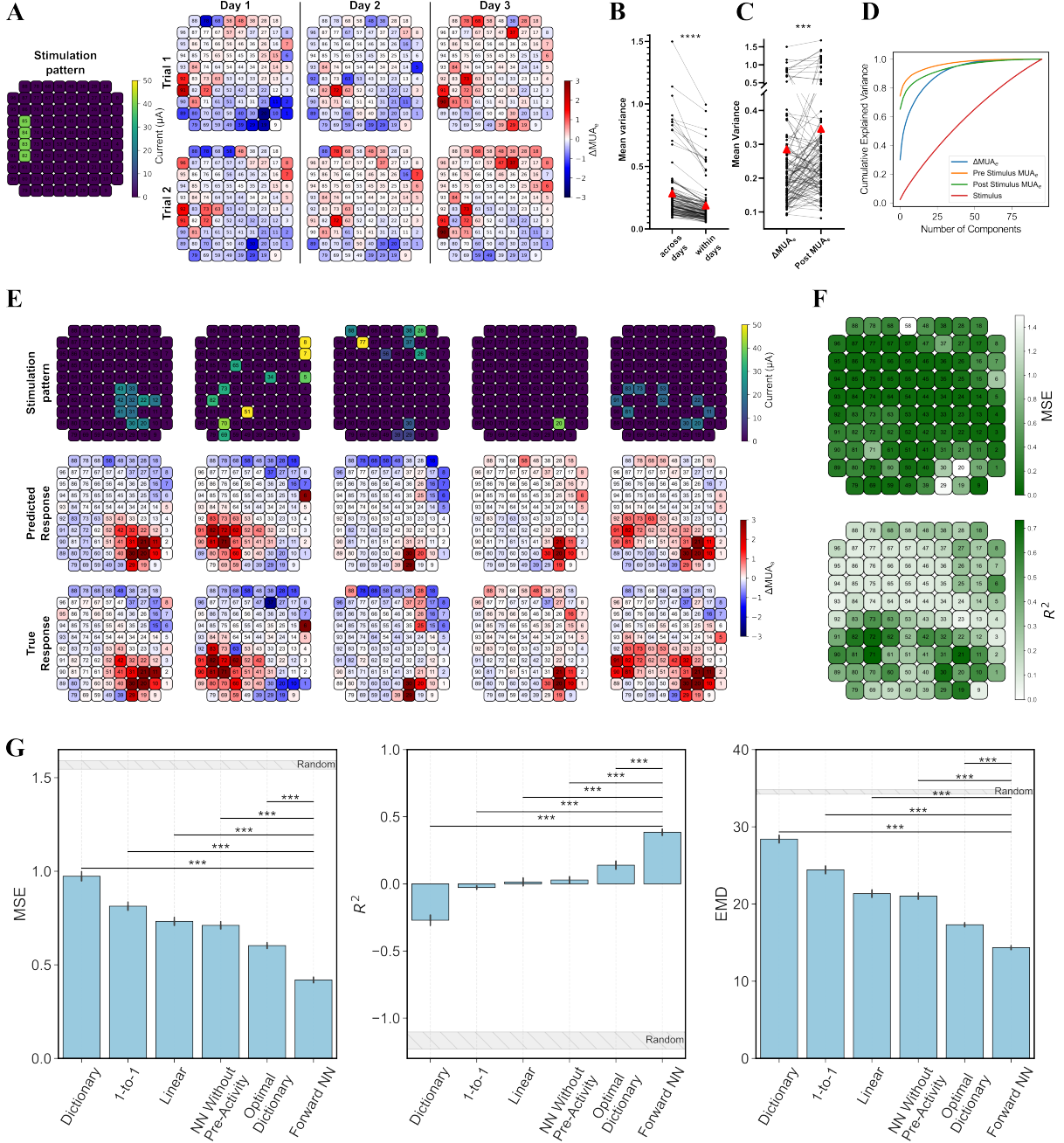


Figure 6.2: **Modeling the neural response to stimulation.** **A)** Example of ΔMUA_e over 3 days to the same stimulation pattern. **B)** and **C)** Comparison of the variance of neural responses to repeated stimulation across days and within days (**B**) and using ΔMUA_e and post-stimulation MUA_e (**C**). **D)** Explained variance as a function of number of principal components. **E)** Forward model predictions and in vivo neural responses on held out samples. **F)** Forward model MSE and R^2 for each channel. **G)** Performance of the forward model in predicting evoked neural responses, measured by MSE, R^2 and Earth Mover's Distance (EMD), compared to baselines.

conventional receptive field model assuming focal activation and thus a one-to-one mapping between stimulation amplitude and neural activity (‘1-to-1’); (2) linear regression with access to average pre-stimulation activity (‘Linear’); (3) a dictionary-based model using an unweighted average of recorded single-electrode responses for each electrode in the multi-electrode stimulation pattern (‘Dictionary’); (4) an optimally weighted combination of single-electrode responses (‘Optimal Dictionary’); (5) the forward NN without access to pre-stimulation activity (NN Without Pre-Activity’); and (6) a random baseline using responses to random stimuli from the training set (Random’), as an estimated lower performance bound.

Performance was evaluated using MSE, R^2 score averaged across electrodes, and an adapted Wasserstein EMD [211], which accounts for the spatial layout of the electrodes (Methods 6.5.5). Only channels which i) elicited spikes on atleast one day and ii) had a variance in neural response across all days greater than a threshold were included in the analysis, resulting in 54/96 valid channels (Methods 6.5.6 and Supplementary Figure 4).

The forward NN model’s predictions aligned well with recorded responses from the test set (Figure 6.2D). Although the MSE was low on most channels, R^2 was higher on the channels near the bottom or upper right of the array, *i.e.* the channels with more spiking activity (Figure 6.2E). Our forward NN model predicted responses with significantly higher R^2 and lower MSE and EMD than all baselines (Figure 6.2F, $p < 0.0001$). The neural network outperformed even the idealized optimal dictionary, providing strong evidence that neural responses from multi-electrode stimulation are not linear combinations of single-electrode responses.

6.3.2 Deep Neural Network-Based Control of Electrically Evoked Activity

While the forward NN enabled accurate prediction of intracortical neural responses, a critical next step is to leverage this understanding to shape neural activity. Previous attempts at electrically evoking targeted neural activity have shown varied success, often limited to simulations [98, 112], retinal explants [95], or prefrontal cortex brain states in non-human primates [113], yet never in awake humans.

To achieve this goal, two deep learning-based methods for inverting the forward neural network model were evaluated. Inspired by the success of gradient-based optimization in controlling populations of neurons via visual stimuli [212, 213], the first approach was a gradient-based method which directly searched the stimulation parameter space (*i.e.* the current to stimulate with on each electrode) for each target response (‘Gradient’). An iterative gradient descent algorithm optimized a stimulation pattern so that, when the stimulus was put through the forward model, the predicted neural response minimized the MSE with the target response.

Gradient descent is computationally expensive and may not be fast enough for real-time deployment. Therefore, the second NN approach consisted of training an inverse neural network (‘Inverse NN’) to map directly from target neural responses to stimuli in a single pass. This inverse network was trained alongside the forward model (with frozen weights), similar to an autoencoder [24, 82, 112]. The network learned to output stimuli that, when input through the forward model, yielded neural responses that minimized MSE with the desired target (Methods 6.5.9).

To evaluate these shaping approaches *in vivo*, optimized stimuli were generated and delivered using the prosthesis, with the participant reporting whether they perceived a phosphene. Below, we compare the performance of these approaches for both natural and

synthetic target responses. Natural target responses consisted of neural activity from the test set of our collected dataset, for which there exists an original ground-truth stimulus that evoked the response on a previous day (Supplementary Figure 5). Synthetic target responses consisted of artificially crafted activity patterns with structures and shapes, potentially infeasible to evoke due to the constraints of the population of neurons under the UEA (Supplementary Figure 6).

The gradient and inverse network approaches were benchmarked against several baselines: (1) the conventional one-to-one mapping (‘1-to-1’), (2) a linear regression model (‘Linear’), and (3) an inverse dictionary approach (‘Dictionary’), where stimulation was computed as a weighted average of the previously recorded stimuli, weighted by target electrode activity (similar to a spike triggered average). For natural targets, the direct replay of stimuli that previously evoked the target responses (‘Replay Stimulus’) was included as a reference measuring day-to-day consistency in neural activity evoked by the same stimulus. A random approach was also included as a lower performance bound, where performance was averaged across responses from random stimuli in the training dataset. The neural activity shaping methods were evaluated using MSE, R^2 score averaged across electrodes, and adapted Wasserstein EMD. Trials were not averaged before computing metrics. See Methods 6.5.9 for full details.

Reproducing Natural Target Responses

For natural target responses, the goal was to reproduce known evoked neural responses on a new day, without knowing the original stimulation pattern. Both the inverse NN and gradient optimization significantly outperformed all baselines in reproducing the target neural responses (Figure 6.3A,C, $p < .05$, more examples in Supplementary Figures 7). Gradient optimization yielded the best overall performance, achieving errors comparable to replaying the original stimuli. Although gradient optimization achieved the best overall

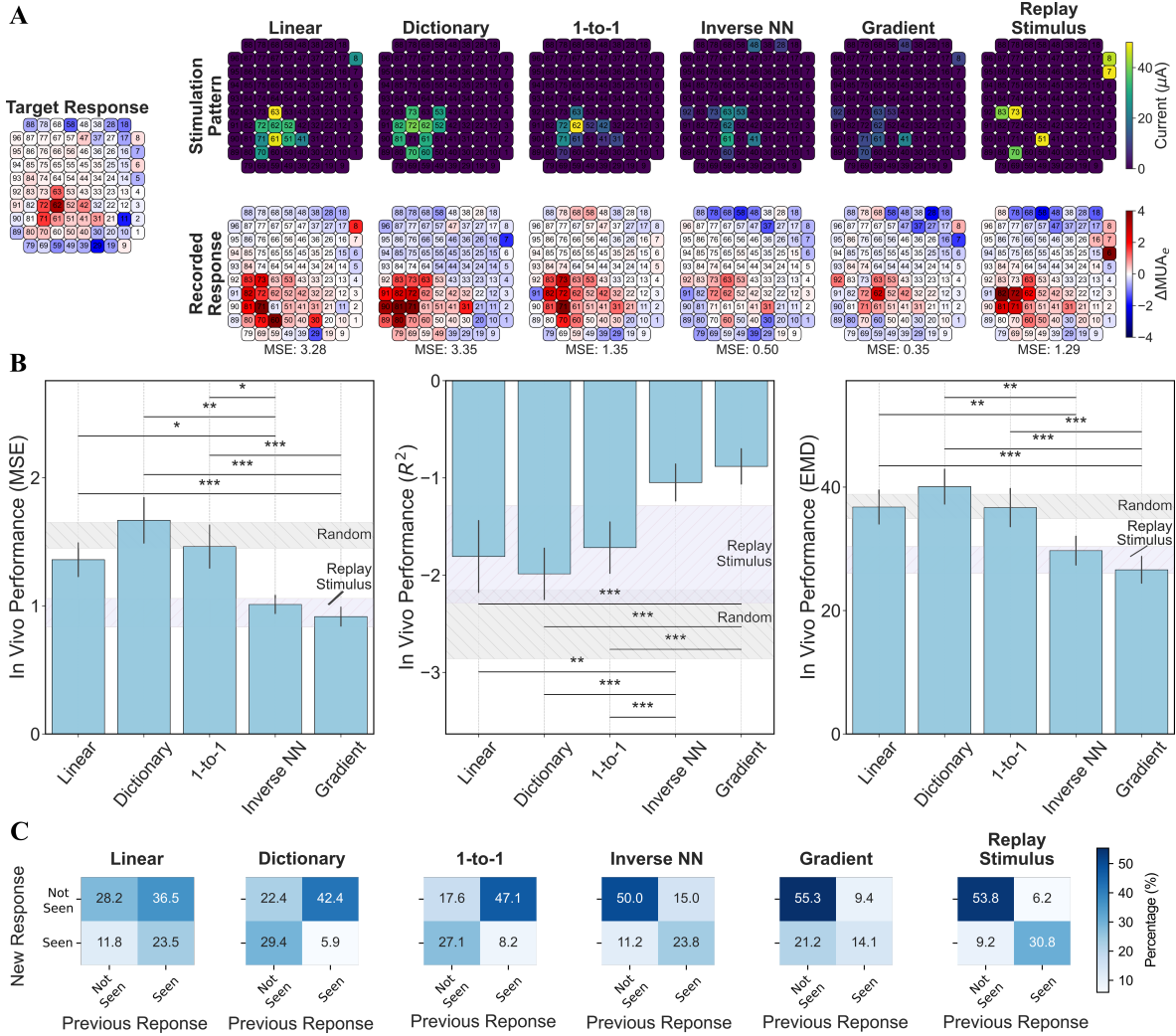


Figure 6.3: Evaluation of neural activity shaping methods. **A)** Optimized stimuli (*top*) and in vivo neural responses (*bottom*) for each shaping method for an example natural target response. **B)** Comparison of errors for each activity shaping method using MSE, R^2 , and EMD. Bars with asterisks denote significant p values: *, **, *** are $p < 0.05$, 0.01 , and 0.001 respectively. Error bars represent SEM. **C)** Confusion matrices showing whether reported perceptions matched the previously observed perceptions associated with the same target neural activities.

results, it required ~ 10 - 20 seconds per example on a desktop with a NVIDIA RTX 3090. In contrast, the inverse NN' was slightly less accurate but inferred stimuli in just ~ 50 microseconds, a difference that could prove critical in real-time applications. Importantly, the perception reported by the participant for DNN-optimized stimulation more closely

matched the participants perception of the original stimuli (detection task, Figure 6.3C).

Generalizing to Synthetic Target Responses

In practical applications, we may wish to generate stimulation patterns for target responses that lie beyond the distribution of recorded neural activity. For example, real-time visual input from a glasses-mounted camera could define a desired percept, but the underlying neural activity required to produce that percept may not exist within the range of previously seen responses. To test the ability of our models to generalize, we used synthetic targets: manually selected structured neural response patterns that may fall outside the feasible neural activity space (Supplementary Figure 6).

Compared to natural targets, synthetic targets had worse MSE but, interestingly, improved R^2 and EMD for all methods (Figure 6.4). However, the relative improvement compared to random stimulation was smaller for synthetic targets than natural, suggesting that neural activity shaping was not as effective as for natural targets. Both the inverse NN and gradient approaches significantly outperformed the baseline methods in shaping neural responses ($p < .05$, Figure 6.4A,B, see Supplementary Figure 8 for more examples). Overall errors were higher for synthetic targets than for natural responses, suggesting constraints imposed by the stimulation paradigm, the training data, or cortical dynamics. Notably, gradient inversion again provided the closest match to the targets, while the inverse NN again showed slightly inferior performance.

Optimized stimulation patterns produced with both gradient-based optimization and the inverse NN achieved target neural responses while requiring lower overall stimulation currents (Figure 6.4C), representing a substantial improvement in stimulation efficiency and safety. To test whether this improvement was simply due to the use of lower amplitudes, we examined the relationship between stimulation current and model performance. Only a weak correlation was observed ($r = 0.25$, Figure 6.4D), indicating that the superior

performance of the DNN-based approaches cannot be explained by current reduction alone.

Instead, model performance was strongly tied to a deeper constraint: the intrinsic structure of the brain’s natural activity manifold. Similar to observations in visually evoked activity [214] and other brain regions [215], electrically evoked responses were confined to a low-dimensional neural manifold, quantified using latent factor analysis [216, 217] (Methods 6.5.10). After accounting for noise, 95% of the manifold variance in evoked responses could be explained with just 10 latent factors, in stark contrast to the 86 factors required to explain 95% of the variance in stimulation parameters. Natural target responses clustered near the manifold, while synthetic targets deviated further (Figure 6.4E, $p < 0.001$, t-test). The farther a target response lay from the neural manifold, the harder it was to control, with a strong correlation between manifold distance and model error ($r = 0.85$, Figure 6.4F). These results suggest that precise control of evoked neural activity may be fundamentally bounded by the brain’s intrinsic response geometry, not simply by stimulation parameters.

6.3.3 Evaluating Simulated Neural Responses to Optimized Stimuli

A critical requirement for stimulus optimization is the ability to accurately predict neural responses before delivering stimulation in vivo. If the forward model can reliably approximate real-world responses, it becomes a powerful tool for designing and refining stimulation strategies without the need for extensive empirical testing. However, discrepancies between simulated and true responses must be quantified to understand the model’s limitations.

To evaluate predictive accuracy, we compared the simulated responses (i.e., those

predicted by the forward model for optimized stimuli) with the true in vivo responses measured in the participant (Figure 6.5A,B).

Across all inversion techniques, simulated responses systematically overestimated the accuracy of neural activity shaping compared to in vivo performance (lower MSE and EMD, higher R^2). However, this overestimation was consistent across methods, and the relative ranking of the different stimulation synthesis techniques was roughly preserved.

Crucially, despite the tendency to overestimate performance, simulated and in vivo MSE were highly correlated across targets ($r = 0.86$ for natural and $r = 0.95$ for synthetic targets; Figure 6.5C). Thus, while absolute errors differed, simulation results reliably predicted which targets and strategies would perform better in vivo, making the forward model a valuable surrogate for optimization.

Further supporting this, we found that even within the simulation, the difficulty of achieving a target response was strongly tied to its distance from the neural manifold. The simulated MSE was highly correlated with the simulated target’s manifold distance (Figure 6.5D), reinforcing that deviations from the brain’s natural activity structure impose fundamental limits on the controllability of neural responses, both in silico and in vivo.

6.3.4 Predicting Perception from Neural Responses

A critical question in the design of stimulation strategies is whether predicting neural activity actually improves perceptual outcomes. One might argue that it would be more direct (and more useful) to predict perception from stimulation parameters alone, bypassing the intermediate step of neural response modeling. To test this, we evaluated whether recorded neural activity provides a better predictor of perception than stimulation parameters themselves.

DNN decoder models were trained to predict the participant’s verbal reports of phosphene detection (seen or not seen, $N = 4416$), brightness (subjective rating 1-5, $N = 427$), and color (4 color categories, $N = 427$). Decoders were trained using different input features: stimulation parameters alone, neural responses alone (ΔMUA_e or ΔMUA_e plus average pre-stimulus MUA_e), or combinations of these inputs. Models were trained via gradient descent to minimize classification error, with performance evaluated using leave-one-day-out cross-validation to assess generalization across recording sessions (see Section 6.5.11 for details).

Neural activity significantly outperformed stimulation parameters alone in predicting detection, color, and brightness ratings (Figure 6.6A-C, $p < 0.05$). When including both pre-stimulus MUA_e and change in envelope multi-unit spiking activity (ΔMUA_e) alongside stimulation parameters (a realistic scenario for future visual prostheses), accuracy improved from 74.2% to 88.7%, corresponding to 56% reduction in error. Similar improvements were observed for color prediction (49.4% to 76.7%, 54% error reduction) and brightness prediction (44.8% to 60.9%, 29% error reduction). F1 scores (Figure 6.6D-F) confirmed these trends.

Together, these results demonstrate that recorded neural activity provides substantially more information about perception than stimulation parameters alone. This finding supports the use of neural response modeling as a critical component of any framework aiming to optimize perceptual outcomes in cortical visual prostheses.

6.4 Discussion

This study demonstrates that deep learning-based stimulus optimization enables robust control of electrically evoked neural activity in the human visual cortex. Using a forward model trained on months of intracortical recordings, we achieved accurate prediction of

neural responses across substantial day-to-day variability, allowing stimulation strategies to be refined *in silico*. Our deep learning-based inverse models successfully shaped neural activity to match target patterns, significantly outperforming conventional methods, while requiring lower stimulation currents and eliciting more consistent perception. Critically, we showed that neural activity, not stimulation parameters alone, provides the strongest predictor of perceptual outcomes, underscoring the central role of neural response modeling for optimizing sensory restoration. While demonstrated here in the context of cortical visual prostheses, this framework provides a generalizable foundation for adaptive, closed-loop control of brain activity across a broad range of bidirectional neural interfaces.

Necessity of Nonlinear Neural Activity Models: When predicting neural responses, our deep NN forward model significantly outperformed multiple linear approaches (Figure 6.2G), providing evidence that neural responses to electrical stimulation are highly nonlinear. This presents a challenge for cortical prostheses, as many traditional stimulus optimization techniques rely on linear receptive field mapping. Our results highlight the necessity of nonlinear, scalable models that can effectively capture the complex relationship between stimulation and neural responses, particularly as implants incorporate more electrodes and higher-density arrays.

This nonlinearity was further reflected in the performance of the inverse models, where deep learning-based approaches significantly outperformed conventional stimulus optimization methods (Figure 6.3 and Figure 6.4). We evaluated two inverse models: (1) an inverse network, which maps target neural responses to stimuli in a single pass, and (2) a gradient-based optimization method, which iteratively refines stimulation parameters to minimize response error.

Both approaches achieved superior neural control compared to 1-to-1 mappings and linear models, with gradient-based optimization yielding the best performance (lower MSE, EMD, and higher R^2). However, the computational cost of ~ 20 seconds per

stimulus limits its current real-time feasibility. In contrast, the inverse network performed inference in ~ 50 microseconds, making it a practical choice for real-time applications. Future implementations may benefit from hybrid approaches, balancing fast inference with periodic fine-tuning to further optimize neural activity shaping.

Activity shaping had smaller performance improvements for synthetic target responses than for natural target responses. It remains an open question whether these discrepancies arise due to fundamental limitations of cortical stimulation (*e.g.*, due to current spread, underlying neural circuitry, or activation of passing axons) or the modeling constraints imposed by generalizing outside the training distribution. If the network were trained on a broader set of response patterns, it is possible that previously unachievable targets could be brought into the distribution of feasible responses, thereby improving generalization. However, the fact that the forward model predicted out-of-distribution stimulus responses reasonably well (Figures 6.2 and 6.5), and the inverse models were trained using the forward model, suggests that at least part of the error might be due to more fundamental constraints on cortical microstimulation.

Notably, we found that electrically evoked neural responses lied on a low-dimensional manifold, and that the distance from a target response to this natural manifold was highly correlated with the error when eliciting that target. While it has been suggested that visually evoked neural activity is distributed according to a lower-dimensional manifold than the stimulus space ([214], but see also [218]), it is more surprising that the same holds for electrically evoked activity. If this is a fundamental constraint due to underlying neural circuitry, then this evoked neural manifold will have important implications for visual prostheses. Future studies will need to determine whether these limitations reflect intrinsic constraints of the neural system or whether they can be mitigated through adaptive training paradigms and expanded stimulation strategies.

Generalization and the Forward Model as a Predictive Tool: A key finding of

this study is the strong correlation between simulated and in vivo neural responses (Figure 6.5), validating the forward model as a predictive tool for designing new stimulation strategies. Even for synthetic targets, the model provided consistent and generalizable approximations of neural responses. The strong correlation between simulated and in vivo responses enables a computationally scalable search for optimal stimulation patterns without requiring extensive in vivo testing. This capability is particularly valuable for refining stimulation protocols and adapting prostheses for individual users.

One challenge in neurostimulation is the variability of neural responses over time, where the same stimulation pattern elicits different responses on different days. This variability likely contributes to the fluctuations in perception reported in prior studies [18]. Our approach partially mitigates this issue by considering the difference in neural activity before and after stimulation (ΔMUA_e), which is more stable across days (Figure 6.2A-C). Additionally, by conditioning stimulation patterns on the average daily pre-stimulation state, our method adapts to variations in latent brain activity. Note that there was no significant performance improvement conditioning instead on per-trial pre-stimulus activity, meaning that accounting for daily drifts in activity was sufficient for shaping neural responses.

Perceptual Relevance and Implications for Visual Prostheses: Our DNN optimized stimuli evoked perceptions more closely matching the perceptions associated with natural target responses (Figure 6.3C), while requiring lower stimulation amplitudes (Figure 6.5C). Additionally, perceptual descriptions were decoded more accurately when including neural responses than from stimulation parameters alone (Figure 6.6), reinforcing the idea that optimizing neural activity—rather than solely stimulation parameters—could enable more precise and reproducible perceptual outcomes.

One question that remains is determining the target neural response associated with a desired perception. Using image neural networks predictive of primate and human V1

could be one solution [98, 212], but our perception decoders might offer another solution. These decoders could be included in the training loop, allowing for perception-driven and neural activity-aware stimulus optimization. This aligns with previous work [206] demonstrating the potential of model-informed microstimulation on modulating perception in sighted primates. By integrating these perception models with our shaping strategy, we could develop fully data-driven approaches for controlling prosthetic vision that exploit the relative abundance of neural activity data, in order to optimize each individual user’s perception.

Limitations and Future Directions: Despite these promising results, several limitations remain. First, our study was conducted in a single participant, and generalization to broader patient populations remains an open question. Second, while our model effectively captured nonlinear neural responses, it does not account for long-term plasticity or adaptive changes in cortical processing over time. Finally, although our inverse models outperformed traditional methods, further improvements are needed to enhance the real-time feasibility of gradient-based approaches.

Future work should explore multi-participant studies, investigate long-term neural plasticity in response to stimulation, and refine hybrid models that combine fast inference with adaptive optimization. Additionally, integrating closed-loop control mechanisms (i.e., where neural activity is continuously monitored and stimulation is dynamically adjusted) could significantly improve the stability and robustness of perception over time.

Towards Closed-Loop Visual Neuroprostheses: These findings help lay the foundation for a closed-loop neural prosthesis, shifting the focus from fixed stimulation patterns to adaptive neural activity shaping guided by predictive models. This has broad implications for sensory brain-computer interfaces, particularly as bidirectional implants (which can both stimulate and record neural activity) become more widespread. By leveraging deep learning models to dynamically adjust stimulation in response to ongoing

neural activity, future prostheses could achieve greater stability, precision, and perceptual fidelity in real-world applications.

6.5 Methods

6.5.1 Electrode Array and Participant

The experiments were conducted as part of a clinical research trial for the development of a cortical visual prosthesis for the blind based on intracortical microelectrodes. The clinical trial was approved by the Clinical Research Committee of the General University Hospital of Elche and registered at ClinicalTrials.gov (NCT02983370). The study adhered to the principles outlined in the Declaration of Helsinki, and all relevant ethical guidelines associated with clinical trial regulation were followed, including EU No. 536/2014 (repealing Directive 2001/20/EC) and EU Commission Directives 2005/28/EC and 2003/94/EC. The participant provided written informed consent prior to their participation. The confidentiality and privacy of the participant were maintained throughout the study, and all procedures were designed to minimize any potential risks to their well-being.

A blind male volunteer (27 years old) with good physical and mental health took part in the clinical trial. He had lost his vision due to a traumatic head injury, after which bilateral enucleation was performed one year prior to his participation.

At the start of the trial, one UEA was implanted into the early visual cortex as an interface for the cortical visual prosthesis [54, 219]. The implant was placed in the right occipital cortex, likely in or near V1 based on cortical anatomy. However, fMRI verification of its precise location was not possible due to the presence of metal fragments in the skull.

The UEA consists of 100 1.5-mm-long electrode shanks (96 of which effectively

connected to the pedestal connector) and can both electrically stimulate the visual cortex to induce visual perception and record the neural activity. A robot-assisted procedure was followed in the UEA implantation [220] in addition to the standard surgical procedure [18]. A platinum wire implanted under the dura was used as reference electrode. After 6 months, the subject underwent another surgery where the UEA was explanted.

During the 6 months of implantation (after which the subject was explanted), the subject performed daily sessions where electrical stimulation of the visual cortex induced visual perceptions (phosphenes). Experimental sessions usually lasted 4 hours per day, 5 days per week. The data collected for this work span across 26 daily sessions, with the last 2 sessions dedicated to the optimized stimulation patterns. The visual perceptions he described were dots or half-circle shapes, and varied in size, color and brightness. Single electrode thresholds were found on 27/96 electrodes. Thresholds ranged from 5 to 90 μA (mean 24 μA , standard deviation 18 μA).

6.5.2 Neural Data Recording and Electrical Stimulation

The Ripple Neuromed Summit Explorer processor was used for both neural recording and electrical stimulation. It was connected to the UEA connector through three Micro2+Stim front-ends (32 channels each). Neural signals were sampled at 30 kHz from 96 channels applying a 0.1 Hz – 7.5 kHz analog filter. The resolution of the analog-to-digital conversion was set to 0.25 μV (16 bit). The stimulation step-size was set to 1 μA . The trigger of electrical stimulation was recorded as a digital event in the stimulation data stream. We used the fast settle option of the processor for all recording channels. With this option, the recording of each channel was blanked out for the duration of each stimulation pulse on any electrode, and an additional 0.5 ms after. Electrical stimulation was controlled through a customized Python interface which allows easy configuration of

the stimulation parameters and timing for each channel.

6.5.3 Data Collection and Task Description

We collected a dataset to train the models to predict the neural response and to find the best stimulation to induce a certain neural activity. Specifically, a set of stimulation parameters was used that has proven effective in inducing visual perception with intracortical electrodes, based on previous literature [18, 205], *i.e.*, a train of 50 pulses lasting 167 ms (300 Hz), where each pulse is rectangular, cathodic first, and charge-balanced (cathodic and anodic duration of $170\mu s$, interphase duration $60\mu s$). For safety reasons, pulses within groups of simultaneously activated electrodes were temporally interleaved to minimize the instantaneous current delivered to the neural tissue (*i.e.*, stimulating during the interpulse gap of other stimulating electrodes, Supplementary Figure 1C). For a stimulation train at 300 Hz and a single-pulse duration of $400\mu s$ there are 2.93 ms between pulses, during which pulses from different electrodes are distributed. We distributed up to six pulses from different electrodes; in case more than six electrodes were included in the stimulation pattern, up to 2 electrodes delivered pulses at the same time.

Three types of stimulation patterns were used:

1. **Random:** A random group of electrodes (between 1 and 10 total) was selected. For each selected electrode, a random stimulation current was chosen between $5\mu A$ to $50\mu A$. Examples of random stimulation patterns are shown in Supplementary Figure 1A. N=5818
2. **Structured:** Occasionally, the patterns were not random but followed structured shapes such as rectangles, triangles, circles or lines. For these structured patterns the current from all the electrodes was set to $15\mu A$ or $45\mu A$. Examples of structured

stimulation patterns are shown in Supplementary Figure 1B. N=484.

3. **Single Electrode:** Additionally, we collected data to determine the current threshold and neural response for each individual electrode, which was then used as a single-electrode dataset to build the forward dictionary models. To find the minimum current needed to induce perception from each electrode, a binary search procedure was used [18]. This led to the collection of a dataset of neural responses to each single electrode stimulation for different currents. Note that this was not included in the training data for the forward and inverse neural network models. N=2389.

In the experiments, the participant was sitting comfortably in a chair, and a sequence of stimulation patterns was delivered to the UEA. After each stimulation, there was a minimum delay of 1 second before the next stimulus. Three tasks were performed, differing in the answers requested from the participant:

1. **No answers needed:** This allowed the collection of the neural response to a large number of stimulation patterns in a short amount of time. N=787.
2. **Detection Task:** The participant reported via keyboard whether the stimulation pattern induced a visual perception or not. N=4811 (25 % of which leading to perception).
3. **Description of the shape, brightness, color, and size of perception:** The participant reported the description by voice, and the experimenter collected the answer in an Excel table. Shape and size were free-form responses which did not vary much throughout data collection, brightness was rated between 0 and 5, and color was a free-form response, which was later grouped into 4 color categories (green, yellow, white, and ambiguous). This protocol took the longest amount of time, so fewer stimulation parameters were tested. N=704.

When using optimized stimulation patterns, task 2 was used, except the patterns of

stimulation were no longer random, but determined by the output of the various stimulation synthesis methods.

6.5.4 Neural Activity Extraction

We calculated the neural activity for each stimulation using MUA_e , an averaged representation of the aggregate spiking activity from several neurons near the electrode tip, as defined in [205]. The raw signal is first filtered between 0.5 and 9 kHz, full-wave rectified, low-pass filtered at 200 Hz, and down-sampled at 1 kHz. Such multi-unit activity approaches are computationally efficient to estimate online and have been shown to provide an accurate representation of neural activity [221] as well as robust performance on decoding tasks [222].

As a measure of the neural response induced by each stimulation pattern, we used the average MUA_e from 100 ms to 200 ms after the end of stimulation, normalized by subtracting the average MUA_e in a baseline window from -110 ms to -10 ms before stimulation (ΔMUA_e). This time window was chosen to be the closest to stimulation offset without any stimulation artifacts, even on the stimulating electrode.

6.5.5 Metrics

We evaluated our models using three complementary metrics. All models were trained using mean squared error (MSE) and additionally evaluated using R^2 and Earth Mover’s Distance (EMD). Note that all metrics were computed on single-trial neural responses without averaging.

For a batch of n samples, the MSE between neural responses $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i \in \mathbb{R}^{N_{elec}}$ was defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2 \quad (6.1)$$

For forward models predicting neural responses, $\mathbf{y}_i$ was an observed neural response, and $\hat{\mathbf{y}}_i$ was a predicted response. For inverse models synthesizing new stimuli, $\mathbf{y}_i$ was the target neural response, and $\hat{\mathbf{y}}_i$ was an observed response to the stimulus synthesized for that target response.

The coefficient of determination (R^2) measures explained variance for each channel. R^2 was calculated for each channel individually and averaged across the array:

$$R^2 = \frac{1}{N_{elecs}} \sum_{j=1}^{N_{elecs}} 1 - \frac{\sum_i (y_{ij} - \hat{y}_{ij})^2}{\sum_i (y_{ij} - \bar{y}_j)^2} \quad (6.2)$$

Here, y_{ij} and $\hat{y}_{ij}$ represent the true and predicted neural response of the j th channel for the i th sample, and $\bar{y}_j$ is the mean of all responses on the j th channel.

We also utilized Earth Mover’s Distance (EMD) to evaluate the proximity of two responses. EMD calculates the cost of “transporting” one sample into another, *i.e.*, treating both neural responses as “piles of dirt”, and calculating the cost of moving the dirt from one response to the next. This metric accounts for the spatial characteristics of the electrode array—it considers responses that are “close” (e.g., off by one electrode) to have a small error—but is computationally expensive for use during training. EMD between two neural responses $\mathbf{y}, \hat{\mathbf{y}} \in \mathbb{R}^{N_{elecs}}$ is defined as:

$$\text{EMD}(\mathbf{y}, \hat{\mathbf{y}}) = \min_{\mathbf{\Gamma} \in G(\mathbf{y}, \hat{\mathbf{y}})} \sum_{i=1}^n \sum_{j=1}^n \Gamma_{ij} d_{ij}, \quad (6.3)$$

where

$$G(\mathbf{y}, \hat{\mathbf{y}}) = \left\{ \mathbf{\Gamma} \in \mathbb{R}^{N_{elecs} \times N_{elecs}} \left| \sum_{j=1}^{N_{elecs}} \Gamma_{ij} = y_i \text{ for all } i, \sum_{i=1}^{N_{elecs}} \Gamma_{ij} = \hat{y}_j \text{ for all } j \right. \right\}.$$

Here, d_{ij} is the physical distance between electrodes i and j , and $G(\mathbf{y}, \hat{\mathbf{y}})$ is the set of valid transport plans between two given recordings [211, 223]. This metric captures both the

spatial relationship between electrodes and the relative magnitudes of neural responses.

6.5.6 Channel Inclusion Criteria

For the evaluation of forward and inverse models, we used only the channels in which any spiking activity was observed (Supplementary Figure 4A shows the electrodes with spiking activity resulting from spike sorting of 9 recordings lasting 5 minutes each in different days) and were ‘responsive’ to stimulation, meaning a variance in the neural response larger than 0.2 (Supplementary Figure 4B shows the variance in the neural response for each electrode). The total number of channels meeting these criteria is 54 out of 96 (Supplementary Figure 4C). These inclusion criteria ensure evaluating our models only in channels that were measuring neural activity.

6.5.7 Statistical Tests

Four different statistical tests were used to assess the significance of our results.

1. Wilcoxon signed-rank tests [224] were used in Section 6.3.1 to show that repeated stimulations had less variance within the same day than across days (Figure 6.2B) and that ΔMUA_e had less variance across repeated stimulations than post-stimulation MUA_e (Figure 6.2C).
2. A one-way ANOVA, with post-hoc independent Student’s t tests for individual differences, was used to test for differences between the forward neural network model and baselines (Section 6.3.1, Figure 6.2F). When controlling for multiple comparisons, we found that individual pairwise tests were not independent, prohibiting family-wise error rate based methods (e.g., Bonferroni), and thus instead corrected for multiple comparisons by controlling the false discovery rate [225].

3. A one-way repeated measures ANOVA was used to test for significant differences between stimulation synthesis methods (Section 6.3.2, Figure 6.3B, Figure 6.4B), with post-hoc paired Student’s t tests, correcting for multiple comparisons by controlling the false discovery rate [225].
4. A one-way ANOVA was used to test for differences across predictor models (Figure 6.6). For comparing each perception model’s performance with the stimulus-only baseline, paired Student’s t tests were used; p-values were corrected for multiple comparisons by controlling the false discovery rate [225].

6.5.8 Forward Model

We developed a deep neural network (‘Forward NN’) to predict the ΔMUA_e elicited from each stimulation pattern. The model is given as input both the stimulation pattern $\mathbf{s}$ (currents on each of 96 electrodes) and the average pre-stimulation activity (-110ms to 10ms) for each electrode on a given day $\bar{\mathbf{y}}_{pre}$, allowing it to account for daily variations in neural responses. The weights θ_1 were trained to minimize the MSE loss:

$$\hat{\mathbf{y}} = F_{\theta_1}(\mathbf{s}, \bar{\mathbf{y}}_{pre}) \quad (6.4)$$

$$\mathcal{L}_{forward} = \frac{1}{n} \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 \quad (6.5)$$

The architecture of the forward model consisted of 10 fully connected layers, with 20% dropout, batch normalization, and residual connections. The stimulation pattern was first input through 4 fully connected layers, with residual connections every 2 layers. This was then concatenated with the average pre-stimulation activity for the day and fed through 6 fully connected layers, with batch normalization and residual connections every 3 layers. ReLU activation functions were applied after each fully connected layer,

with the final layer outputting 96 units corresponding to the predicted ΔMUA_e for each electrode. In total, the model had approximately 1.2 million trainable parameters.

The model was trained to minimize MSE between predicted and observed responses. The Adam optimizer was used with a learning rate of 0.0001 and a batch size of 256 for model training. A scheduler was used to decrease the learning rate by a factor of 0.5 every 200 epochs. Early stopping was also used to terminate training early based on a validation set performance, resulting in a total of 911 training epochs.

The ‘Forward NN’ was compared against several baselines:

Forward without pre-activity: To test the effect of ablating the average pre-stimulation activity, we trained an otherwise identical neural network to the forward model, but with only the stimulus as input.

1-to-1: Each stimulating electrode was assumed to focally activate only the nearby neurons. Thus, the predicted response on each channel was directly proportional to the stimulating current on that electrode. The maximum observed neural activity was assumed to be proportional to the maximally delivered current ($50\mu A$)

Dictionary: Loosely inspired by [95], this method predicts multi-electrode stimulation responses via a lookup ‘dictionary’ of responses to single-electrode stimuli. Given an input multi-electrode stimulation $\mathbf{s} \in \mathbb{R}^{N_{elec}}$ and selected channel $i \in [0, N_{elec}]$, then $\mathcal{D}\{\mathbf{s}, i\}$ is a dictionary function which outputs the average response in the training dataset across single-electrode stimuli on electrode i with the nearest amplitude to s_i (*e.g.* if s_2 was $40\mu A$, and the training set had 5 responses to single-electrode stimuli on electrode 2 at both 45 and $25\mu A$, then $\mathcal{D}\{\mathbf{s}, 2\}$ would return the average of the 5 $45\mu A$ responses). We evaluated two dictionary variants. First, the neural responses for each individual single-electrode stimulus were averaged to get the predicted multi-electrode response (‘Dictionary’): $\hat{\mathbf{y}}_{avg} = \frac{1}{|\mathbf{s}_{>0}|} \sum_{s_i \in \mathbf{s}_{>0}} \mathcal{D}\{\mathbf{s}, i\}$; and second, the ideal linear combination of the single electrode responses was computed (‘Optimal Dictionary’): $\hat{\mathbf{y}}_{opt} = \sum_{s_i \in \mathbf{s}_{>0}} w_i \cdot \mathcal{D}\{\mathbf{s}, i\}$

where w_i are learned weights optimized for each given stimulus and its true neural response $\mathbf{y}$. This idealized dictionary would be infeasible to deploy in a real device, but gives a baseline of the maximum performance assuming linear combination of single-electrode responses.

Linear: A linear regression model was trained on the same dataset as the deep neural network. This model was given both the stimulus vector and the pre-activity to enable a fair comparison to the neural network forward model. The model can be formulated as: $\hat{\mathbf{y}}_{linear} = \mathbf{W}\mathbf{s} + \mathbf{b}$, where $\mathbf{s}$ is the stimulus concatenated with the average pre-stimulation activity for that day, $\mathbf{W}$ is a weight matrix and $\mathbf{b}$ is a bias term.

Random: This approach randomly paired stimuli with neural activities from the training set, and averaged across 500 random pairings, serving as a lower performance bound.

6.5.9 Inverse Models

We developed several inversion methods using the forward model to target particular neural activity patterns. We compare our inversion methods with several baseline and state-of-the-art inversion methods.

Gradient: We developed a gradient-based optimization method, which freezes the training parameters of the forward model and uses back propagation to maximize the input stimulation pattern for a desired target neural response. This method directly optimizes a stimulation pattern $\mathbf{s}$ to minimize the difference between the forward model’s predicted response $F(\mathbf{s})$ and the target response $\mathbf{y}$, with L1 regularization on the stimulus to minimize stimulation amplitudes and the number of active electrodes:

$$\mathcal{L}_{\text{grad}}(\mathbf{s}) = ||F(\mathbf{s}, \bar{\mathbf{y}}_{pre}) - \mathbf{y}||_2^2 + \lambda \cdot ||\mathbf{s}||_1 \quad (6.6)$$

$$\mathbf{s}_{grad} = \underset{\mathbf{s}}{\operatorname{argmin}} \mathcal{L}_{grad}(\mathbf{s}) \quad (6.7)$$

Where again $\bar{\mathbf{y}}_{pre}$ is the average pre-stimulation activity for the given day. We employed the Adam optimizer with learning rate of 0.001 and ran optimization for 10000 iterations or until convergence.

Inverse NN: We also built an inverse neural network model to directly map from stimuli to target neural responses. In this method, the weights of the forward model were frozen, and a secondary neural network was trained together with the frozen forward model in an autoencoder fashion, to approximate the forward model’s inverse function F^{-1} , with weights θ_2 . The architecture was the same as our forward model with three fully-connected layers of 96 units each, residual connections every other layer, and ReLU activations throughout the network, with dropout layers (rate=0.3) applied after each hidden layer to prevent overfitting. The inverse network also took as input the average pre-stimulation activity $\bar{\mathbf{y}}_{pre}$ to account for daily variations in neural activity. The networks output was scaled by a factor α to ensure generated stimulation patterns fell within the valid range (0-50 μA). The weights of the network θ_2 were trained to minimize:

$$\begin{aligned} \mathbf{s}_{inv} &= F_{\theta_2}^{-1}(\mathbf{y}, \bar{\mathbf{y}}_{pre}) \\ \mathcal{L}_{inv} &= ||F(\alpha \cdot \mathbf{s}_{inv}, \bar{\mathbf{y}}_{pre}) - \mathbf{y}||_2^2 + \lambda \cdot ||\mathbf{s}_{inv}||_1 \end{aligned}$$

where F represents the frozen forward model, F^{-1} is our inverse network, and $\mathbf{y}$ is the

target neural response pattern. With this loss, the inverse network is trained to generate stimulation patterns that, when passed through the forward model, reproduce the target response.

1 to 1 mapping: For our baseline comparisons we first implemented a conventional approach that assumes a direct relationship between stimulation and neural response. Here, stimulating amplitude was directly proportional to the desired activity on the underlying channel. Thus, there was a 1-1 mapping between target activations on each channel and stimulating currents on the same electrode. A target activity of 0 would have a current of 0, and a target activity of 4 (the maximum among our targets) would have a stimulating amplitude of $50\mu A$

Dictionary: We evaluated an inverse dictionary approach, which leverages the dictionary ($\mathcal{D}$) of recorded neural responses ($\tilde{\mathbf{y}}$) and stimulation pattern ($\tilde{\mathbf{s}}$) pairs to optimize stimulation patterns for a new target $\mathbf{y}$. For a given target activation pattern, the dictionary stimulus was defined as the weighted average of every stimulus in the dictionary, weighted by the mean recorded response on channels active in the target response, similar to a traditional spike-triggered average.

$$\tilde{w} = \frac{1}{|y > 0|} \sum_{y_i \in \mathbf{y} > 0} \tilde{y}_i \quad \forall \tilde{\mathbf{y}}, \tilde{\mathbf{s}} \in \mathcal{D} \quad (6.8)$$

$$\mathbf{s}_{dictionary} = \frac{1}{\sum_{\tilde{\mathbf{y}}, \tilde{\mathbf{s}}, \tilde{w} \in \mathcal{D}} \tilde{w}} \sum_{\tilde{\mathbf{y}}, \tilde{\mathbf{s}}, \tilde{w} \in \mathcal{D}} \tilde{w} \cdot \tilde{\mathbf{s}} \quad (6.9)$$

Linear: We also evaluated a linear regression model trained to map from neural responses back to stimulation patterns:

$$\begin{aligned} \mathbf{y} &= \mathbf{W}\mathbf{s} + \mathbf{b} \\ \mathbf{s}_{linear} &= \mathbf{W}^\dagger(\mathbf{y} - \mathbf{b}) \end{aligned} \quad (6.10)$$

where W and b are learned parameters optimized to minimize MSE on the training data and $W^\dagger$ is the pseudo inverse.

Replay Stimulus: We evaluated the activity generated by replaying the stimulation that evoked the target activity for Natural targets. This approach serves as an upper performance bound (the best achievable performance is limited by the variability of the response to the same stimulus).

Random: The performance of a random stimulation synthesis strategy for each target response was calculated using the average distance (measured using each metric) from each target to each response to random stimuli in the forward model training dataset.

For safety constraints, for all methods we only select the ten electrodes of $\mathbf{s}$ with the largest current amplitude.

6.5.10 Latent Factor Analysis

To quantify the manifold underlying neural activity we used latent factor analysis [216], which has been widely used previously (*e.g.* [217]). Under this model, neural responses $\mathbf{y} \in \mathcal{R}^{N_{elec}}$ are assumed to be distributed according to a number of latent factors $\mathbf{z} \in \mathcal{R}^D; \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$:

$$\mathbf{y}|\mathbf{z} \sim \mathcal{N}(\mathbf{\Lambda}\mathbf{z} + \boldsymbol{\mu}, \boldsymbol{\Psi}) \quad (6.11)$$

where $\boldsymbol{\mu} \in \mathcal{R}^{N_{elec}}$ is the mean response for each channel, $\boldsymbol{\Psi} \in \mathcal{R}^{N_{elec} \times N_{elec}}$ is a diagonal covariance matrix, and $\mathbf{\Lambda} \in \mathcal{R}^{N_{elec} \times D}$ contains loadings mapping latent factors back to the neural activity.

$\mathbf{\Lambda}$ and $\boldsymbol{\Psi}$ were estimated using expectation-maximization with the python package scikit-learn [226], fit to the same dataset of neural responses used to fit the forward model. We chose $D = 10$, which captured 95% of the manifold variance in the observed responses (*i.e.*, after accounting for noise). Similar to [217], the neural manifold was defined as

the column space of $\mathbf{\Lambda}$. The reported manifold distance was calculated as the Euclidean distance from each neural response to its orthogonal projection onto the columns of $\mathbf{\Lambda}$ (Equation 6.12). Note that even for the observed neural responses themselves, there is still a nonzero distance to the underlying manifold due to noise and variance not captured in the 10 components (Figure 6.4E, green dotted line). Natural targets had similar manifold distances as the neural manifold responses, and synthetic targets had significantly larger manifold distances.

$$d_m(\mathbf{y}) = \|\mathbf{y} - \mathbf{\Lambda}(\mathbf{\Lambda}\mathbf{\Lambda}^T)^{-1}\mathbf{\Lambda}^T\mathbf{y}\|_2 \quad (6.12)$$

6.5.11 Predicting Perception

To validate the relevance of neural responses to perceptual measurements, we trained a decoder to predict verbal perceptual responses from recorded activity (descriptions of phosphene detection, color, and brightness rating). See Section 6.5.3 for more details on the task description.

The perception model used five fully connected layers of 256 units each with residual connections to every other layer; batch norm layers, dropout layers, and ReLU activations after each hidden layer. Training was performed using the Adam optimizer and cross-entropy loss. For each model type, dropout, batch size, ΔMUA_e window, dropout rates, and learning rate were tuned independently for each model configuration, and number of parameters was balanced across model types, to ensure fair comparison. Models were tested using leave-one-day-out cross validation to evaluate generalization across sessions.

The model receives as input neural data, either pre- and post-stimulation ΔMUA_e , or the stimulation amplitudes per electrode. The model outputs were categorical variables corresponding to the perceptual attributes: binary classification for phosphene detection (seen/not seen), multi-class classification for color perception, and ordinal classification

for brightness ratings.

6.6 Data and Code Availability

The data and code used for this study will be made available at https://github.com/jgranley/deep_neural_shaping

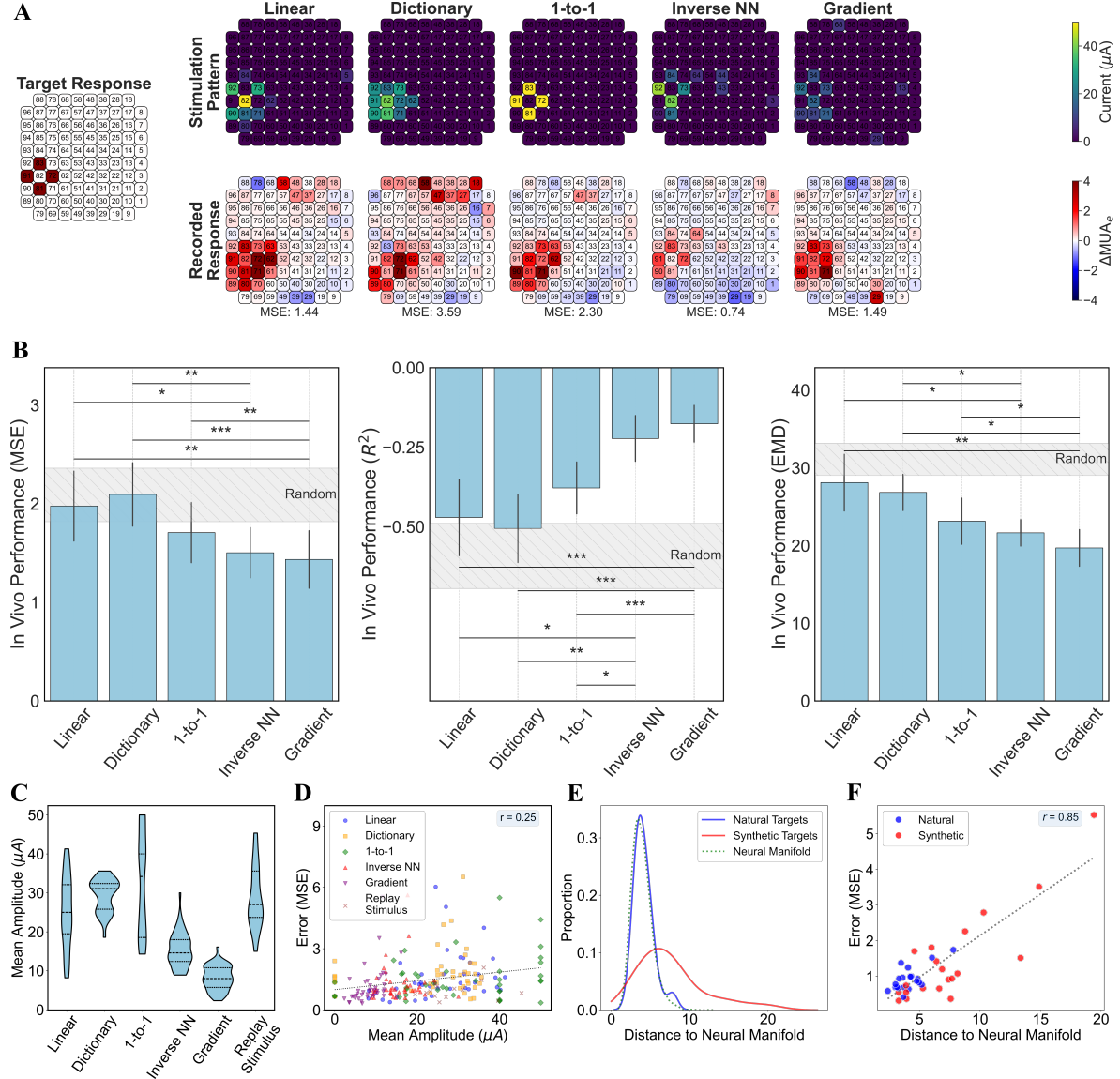


Figure 6.4: **Evaluation of neural activity shaping methods for synthetic targets.**

A) Optimized stimuli (*top*) and in vivo neural responses (*bottom*) for each activity shaping method for an example synthetic target response. **B)** Comparison of errors for each method using MSE, R^2 , and Wasserstein EMD. Bars with asterisks denote significant p values: *, **, *** are $p < 0.05$, 0.01, and 0.001 respectively. Error bars represent SEM. **C)** Average stimulation amplitude for each method across natural and synthetic targets. **D)** Amplitude was only slightly correlated with neural response reconstruction error. **E)** Kernel density estimate plot showing the distance to the low dimensional ($D=10$) natural neural response manifold across target neural responses. The manifold was quantified using latent factor analysis. Synthetic targets lied further from this manifold than natural targets. **F)** Target responses which were further from the latent manifold had larger reconstruction error (shown for ‘Gradient’ responses).

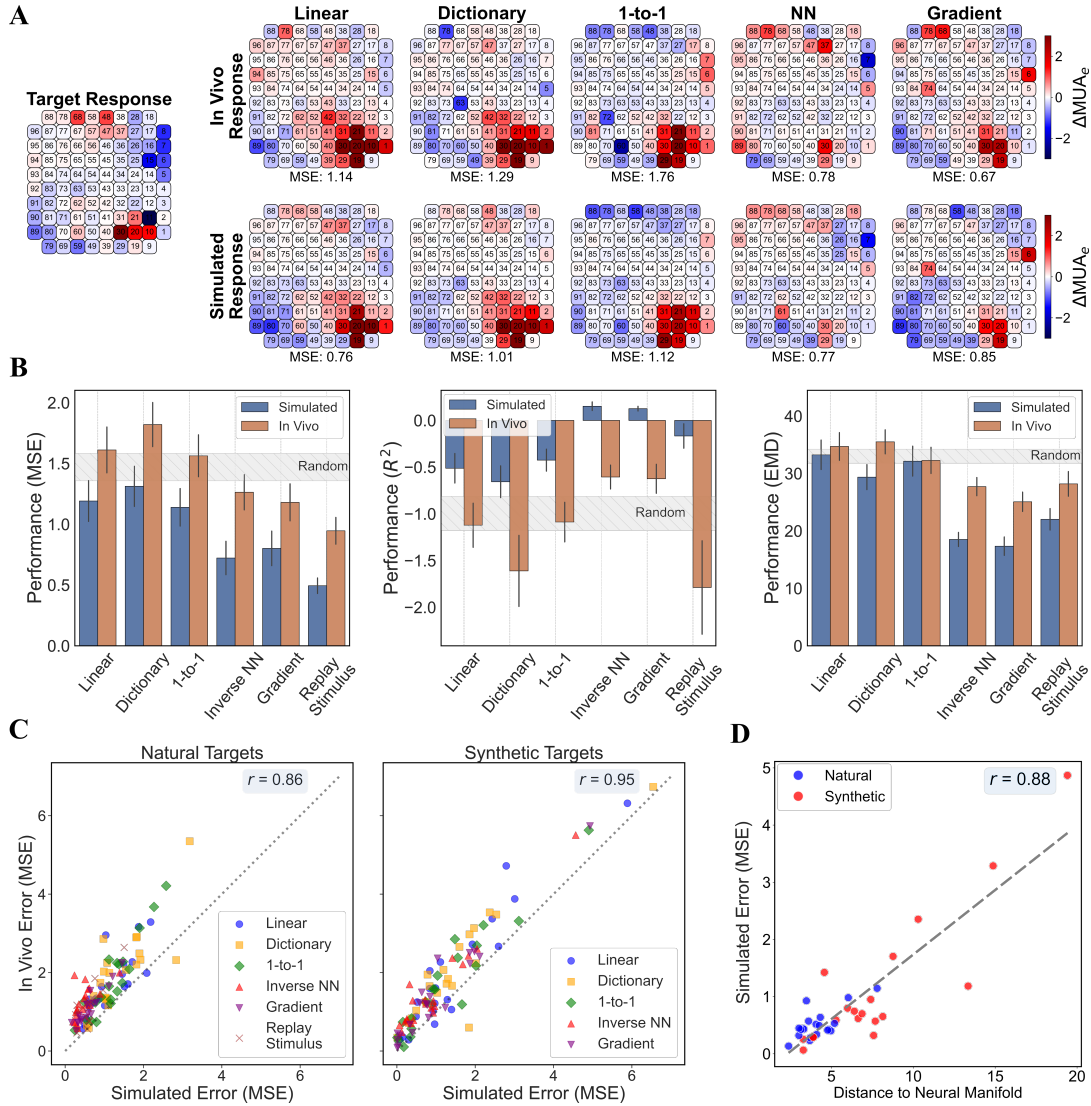


Figure 6.5: **Generalization of neural activity shaping strategies from simulated models to in-vivo measurements.** **A)** Simulated neural responses (*bottom*) and *in vivo* responses (*top*) for an example target. Simulated responses were obtained by running the synthesized stimulus through the forward neural network model. **B)** Comparison of errors for simulated and in vivo responses across methods. **C)** Simulated vs in vivo neural response reconstruction error (MSE) for each shaping method, for natural (*left*) and synthetic (*right*) targets. **D)** Target responses with a larger distance to the latent neural manifold also had larger simulated errors.

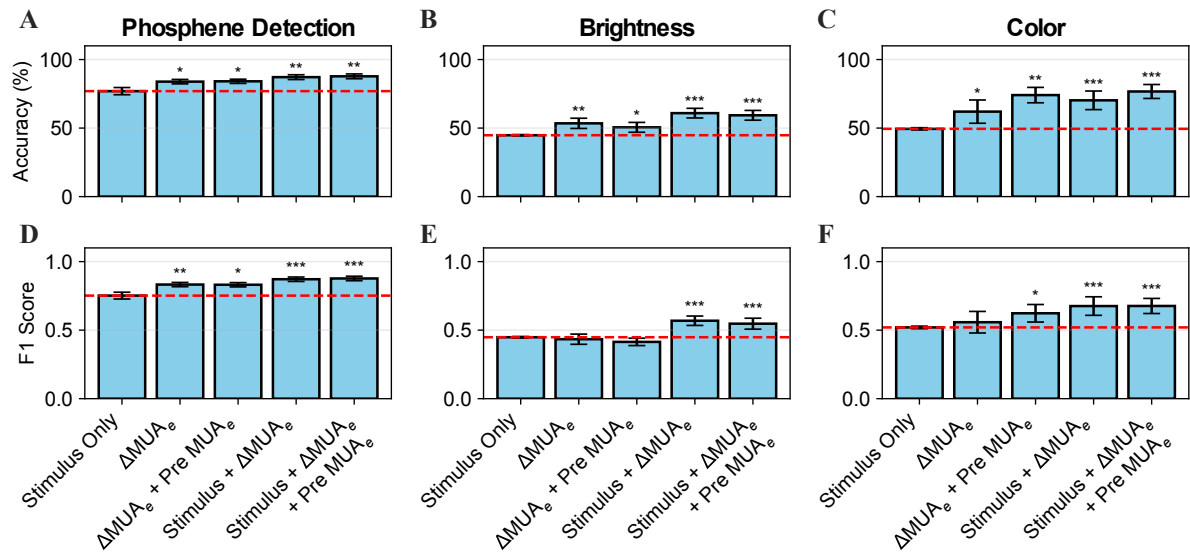


Figure 6.6: **Perceptual relevance of evoked activity patterns.** **A-C)** Accuracy of decoder models for phosphene detection ($N = 4416$), brightness ($N = 427$), and color ($N = 427$) perception, decoding using different input features. Leave-one-day-out cross validation was used, and performance aggregated across predictions from different folds (mean $\pm$ SEM). Asterisks denote significant differences when compared to stimulus only: *, **, *** are $p < 0.05$, 0.01 , and 0.001 respectively. **D-F)** F1 scores for the same perceptual attributes and model inputs as A-C.

6.7 Acknowledgments

This work was financially supported by:

Ministerio de Ciencia, Innovación y Universidades Grant DTS19/00175.

Ministerio de Ciencia, Innovación y Universidades Grant PDC2022-133952-100.

European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No. 899287 (NeuraViPeR).

European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 861423 (enTRAIN Vision).

Innovative Neurotechnology for Society (INTENSE). Dutch Neurotechnology Consortium.

Generalitat Valenciana, Directorate General of Science and Research, PROMETEO Grant CIPROM/2023/25.

Swiss National Science Foundation projects CA-DNNEdge (208227)

National Library of Medicine of the National Institutes of Health under Award Number DP2-LM014268 to MB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Chapter 7

Conclusion and Future Directions

Visual prostheses show great promise for treating blindness, but current devices have failed to meet consumer expectations. One reason for this is due to naive stimulation strategies which do not elicit prosthetic vision of high enough quality to be useful for everyday tasks.

To help address this, in this dissertation we made four main contributions towards our goal of model-informed and adaptable stimulation strategies:

1. Psychophysically validated phosphene models for epiretinal prostheses.

We integrated decades of patient data into a neuroanatomically inspired predictive model which on average decreased prediction errors by 88%, and is the current state of the art model for epiretinal devices (Chapter 3 and Section 5.5.1).

2. A deep learning framework for model-informed stimulus encoders.

We designed a general framework where a deep neural network is trained to invert a phosphene model, which dramatically improved stimulus encodings in simulated patients. Similar approaches have become widespread in visual prostheses research (e.g. [82, 99, 101–109]).

- 3. Human-in-the-loop optimization for deep stimulus encoders.** We designed a preferential Bayesian optimization framework to personalize deep stimulus encoders based on user feedback, which we showed produced near-optimal percepts for simulated users, was robust to noise and erroneous modeling assumptions, and generalized to humans viewing simulated prosthetic vision.
- 4. First successful in-human neural activity shaping via deep learning.** We deployed our encoder-decoder framework to shape electrically evoked neural activity towards desired activity patterns in a human with an intracortical prosthesis.

Together, these are important steps towards enabling smarter, AI-endowed closed-loop visual prostheses which intelligently choose stimulation parameters according to an accurate model of perception, adaptive each individual user’s subjective preferences, and modulated by the current cortical state. Hopefully, this will help enable more effective visual restoration and improve the functional utility of current and future visual prostheses.

Safety. Since it is difficult to predict the behavior of deep learning models, using a deep stimulus encoder in real patients without manual oversight could raise safety concerns. It may be possible for a deep encoder to produce unconventional stimuli, potentially leading to adverse effects. Most devices come with firmware responsible for ensuring stimuli stay within FDA-approved safety limits. However, incorporating safety constraints into the training procedure for stimulus encoders is an important next step for building safe and reliable stimulation strategies.

Broader Implications. Although the work here was presented in the context of visual prostheses, our encoder-decoder framework for stimulus encoding is general, and may be applicable to a variety of stimulating brain-computer interfaces. In general, it may be applicable whenever the stimulus space is large, and there exists a forward model

mapping stimuli to responses. Such forward models (*e.g.*, [150, 156, 157, 160]) and even deep stimulus encoders (*e.g.*, [114, 115, 195]) have now been developed across a variety of sensory modalities.

Limitations and Future Work. One limitation of the methods presented here is that they do not account for the temporal effects of stimulation such as phosphene fading. Recently, there have been developments in accurate modeling of temporal effects [80, 99] which could be included in a deep stimulus encoder framework. Additionally, although it has been widely used in simulation, the encoder-decoder framework presented here has only been validated in one human prosthesis recipient via our work on neural activity shaping. Future work should help bridge the gap towards practical applicability to deploy and evaluate deep stimulus encoders in humans.

Appendix A

Chapter 4 Appendix

A.1 Surrogate Model

This section covers specific implementation details about the surrogate model as well as observations on its performance.

A.1.1 Implementation Details

Dataset To create training data for the surrogate model $\hat{f}_\phi$, we used the phosphene model described in [22] and implemented in pulse2percept v0.8 [74]. 50,000 stimuli were created by first selecting a number of electrodes to stimulate between 1 and 30 randomly chosen electrodes, then randomly selecting an amplitude between 1 and 10 (specified as a multiple of the assumed threshold current) and frequency between 1 and 200 Hz for each electrode. In addition, between 10 and 100 electrodes were chosen to act as “noise” electrodes, where either amplitude or frequency was given a nonzero value, but not both. The purpose of these electrodes was for the surrogate model to learn that both a nonzero amplitude and a nonzero frequency are required to produce a visible percept. We used an 80-20 train-test split. As the surrogate model is highly dependent on patient-specific

parameters ϕ , we generated new data and fit a separate surrogate for each of the following ϕ : $((\rho, \lambda) \in \{(150, 100), (150, 1500), (800, 100), (800, 1500)\})$.

Network Architecture The surrogate model $\hat{f}_\phi$ used a fully-connected architecture. The input to the model was a stimulus matrix $\mathbf{s} \in \mathbb{R}_{\geq 0}^{n_e \times 3}$, which was identical to the input to f . The stimulus matrix was split into amplitude and frequency components (pulse duration was not used due to poor model performance), which were fed through a FC layer. The outputs of both FC layers were concatenated and fed through another FC layer. Concurrently, the model computed the element-wise product of the amplitude and frequency components and passed it through a separate FC layer. The outputs of the previous two layers were then concatenated and fed through a final FC layer with output size 49×49 .

The model was trained for 45 epochs using AdamW [227] optimizer and MAE loss.

A.1.2 Approximating the Forward Model

The surrogate model was able to accurately approximate the true phosphene model f . Table A.1 shows MAE over the validation set (10,000 percepts) for all 4 trained $\hat{f}_\phi$. Visually, the predicted percepts were nearly identical to the ground truth.

Table A.1: Surrogate model performance

ϕ	$\rho = 150 \ \lambda = 100$	$\rho = 150 \ \lambda = 1500$	$\rho = 800 \ \lambda = 100$	$\rho = 800 \ \lambda = 1500$
MAE	0.0119	0.0189	0.0078	0.0115

A.1.3 Predicted Stimuli

Despite the low surrogate validation error, training with the surrogate model would often result in the encoder suggesting almost adversarial stimuli; that is, stimuli that if

fed through the true forward model f would lead to drastically different percepts than if fed through the surrogate model $\hat{f}$ (see Fig. A.1.1). With these adversarial-like stimuli, the encoder appears to be performing well under the surrogate model, but performs poorly when the same stimuli are input to the true forward model. We identify this as the primary disadvantage of using a surrogate model and resolving this issue remains an open research problem for end-to-end training with surrogate methods.

We noticed several issues caused by the effects of varying stimulus parameters on phosphene appearance. For example, increasing amplitude increases size and brightness, while increasing frequency increases brightness only. We noticed a larger mismatch between the surrogate and the forward model on the extreme ends of the spectrum (*e.g.* very high frequency, low amplitude), resulting in the encoder settling into a minimum that does not exist in the true forward model. It is important to note this disparity appears despite a high training accuracy of the surrogate alone. Although these examples are specific to the bionic vision application, we expect surrogate models derived to describe other neuromodulation technologies to suffer from similar limitations.

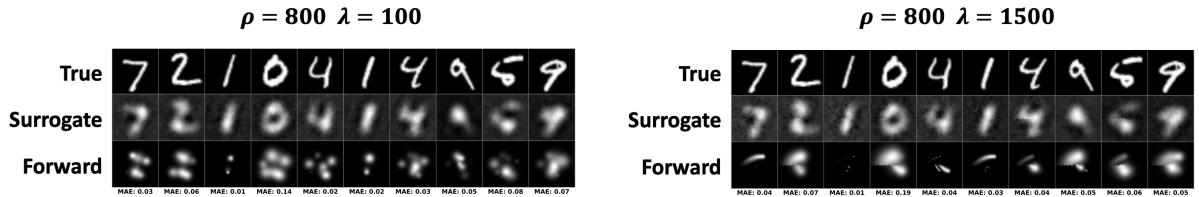


Figure A.1.1: The encoder would often suggest stimuli that lead to drastically different percepts when fed through the surrogate model ($\hat{f}$, middle row) as compared to the true forward model (f , bottom row). Examples are shown for $\rho = 800$; $\lambda = 100$ (*left*) and $\rho = 800$ $\lambda = 1500$ (*right*).

A.2 Hyperparameter Selection

In this section, we detail how HNA hyperparameters (l , k , α , and β) were chosen.

VGG Loss To choose the layer of the VGG network to use for VGG loss (l) we performed cross validation across a set of candidate layers. Previous studies [163] have shown that the first layer with each of the 5 convolutional blocks perform well for neural style transfer. Thus, we choose these as our candidate layers. For cross validation, we trained HNA for 50 epochs using each candidate layer. The resulting phosphenes are shown in Figure A.2.1. Using earlier layers, the VGG term performs similarly to MAE, and phosphenes are disconnected. We chose layer 5_1 based on its perceived ability to capture high-level perceptual differences between images, although layer 4_1 also performs similarly.



Figure A.2.1: Phosphenes produced by HNA encoder with different layers chosen for VGG loss. Layer 5_1 denotes the first layer within the fifth convolutional block.

Laplacian Smoothing We chose to use a kernel size 5 for the Laplacian filter used to estimate the second derivative (k , Eq. 4.6). The size of the filter controls the scale on which smoothing is applied (*i.e.*, smaller filters sizes only encourage continuity within a small local region, whereas larger filters encourage continuity within a larger region). Size

5 was chosen because larger filters were observed to over-smooth the image, while smaller filters still led to highly disconnected phosphenes.

Joint Perceptual Metric We performed cross validation to find the best values for α and β . Instead of using one value, we found scheduled weighting to be crucial for performance. The scheduler incrementally increased the weight of the VGG loss (β) from 0 while simultaneously decreasing the initially high weight on the smoothing constraint (α). This was motivated by the observation that the VGG loss performed poorly during early iterations when the predicted phosphene was near-random.

Under this scheduled weighting strategy, the loss is dominated early on by the MAE and smoothing terms. This encourages the the model to just output reasonable encodings. As training progresses, the predicted phosphenes become higher quality, causing the VGG loss to perform better, and thus the smoothing term is no longer as important.

Additionally, we found it beneficial to temporarily decrease the learning rate by a factor of 10 for a short 'warm-up' duration following each increase in β , before resetting to 50% of the prior learning rate. This results in the learning rate gradually decreasing throughout training by a factor of around 100. Throughout the paper, we use $\alpha = 0$ and $\beta = 0.00008$ for comparisons of loss values.

A.3 COCO Dataset

For the COCO task (Section 4.6.2), we used subset of images from the MS-COCO dataset [166]. MS-COCO was chosen due to its selection of common household objects relevant to the daily life of prosthesis users, as well as availability of ground-truth segmentation masks. To select the images suitable for prosthetic vision, we filtered out images according to the following criteria:

1. **Too cluttered.** Any image with greater than 15 total objects was removed. Removed: 15566
2. **Select chosen objects.** Any image that did not have at least 1 object from the selected categories that was larger than 4% of the total image was removed. Removed: 42289
3. **Too many.** Any image with greater than 5 objects meeting criteria 2 was removed. Removed: 1017
4. **Too dim.** Any objects in the image with average pixel brightness less than 50 were discarded. If this resulted in an image having 0 remaining objects, the image was removed. Removed: 434

This resulted in a total of 47,532 training images and 11,883 test images (80-20 train-test split). The objects in the remaining images were segmented out using the ground-truth segmentation masks, resized to (49, 49), and converted to grayscale. The distributions of classes used is shown in Figure A.3.1.

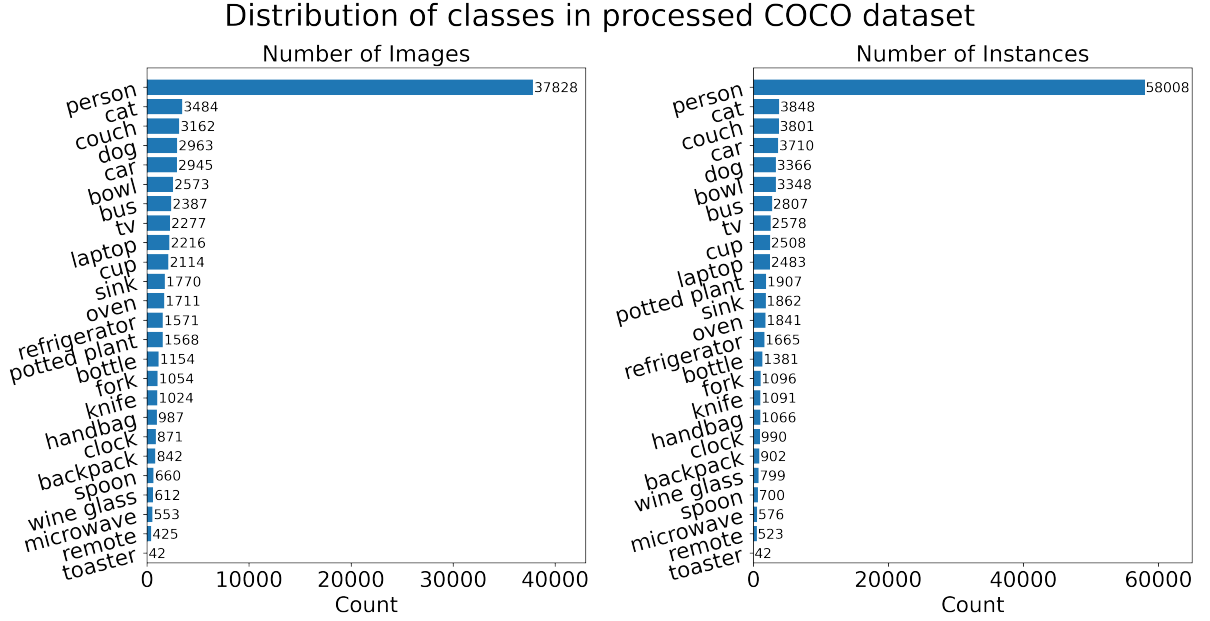


Figure A.3.1: Number of images (*left*) or instances (*right*) of each category in the processed COCO dataset.

A.4 Predicted Stimuli

Here, we directly examine the stimuli resulting from HNA and naive encoders. Stimuli and their resulting phosphenes for example images from the test set are shown in Figure A.4.1. The naive encoder produces stimuli with constant frequency (20 Hz) and pulse duration (0.45 ms), which are not shown.

We make the following observations about the predicted stimuli:

- Both encoders activate electrodes corresponding to the shape of the target image. In naive stimuli, the amplitude directly corresponds to the pixel brightness. In HNA stimuli, the distributions of amplitude, frequency, and pulse duration across the electrodes is more complex and harder to characterize, but lead to higher-quality phosphenes.
- HNA uses amplitudes inversely proportional to ρ .
- For small ρ , HNA primarily uses amplitude to control brightness. For large ρ , HNA

primarily uses frequency to modulate brightness, keeping amplitudes low to limit phosphene size.

- HNA uses small pulse durations to create lines parallel to the underlying axon NFB (*i.e.*, it utilizes the streaked phosphenes to its advantage), and large pulse durations to create lines perpendicular to the underlying NFB. In other words, HNA was able to exploit application-specific (*i.e.*, neuroanatomical) information that is baked into the forward model.
- On average, HNA uses more electrodes, larger frequencies and pulse durations, and smaller amplitudes than the naive encoder. A large active electrode count and high pulse durations may not be desirable for some prostheses, due to tissue activation and frame rate limits. We found that it was easy to constrain these parameters using regularization on the output stimuli, at the cost of slightly decreased performance.

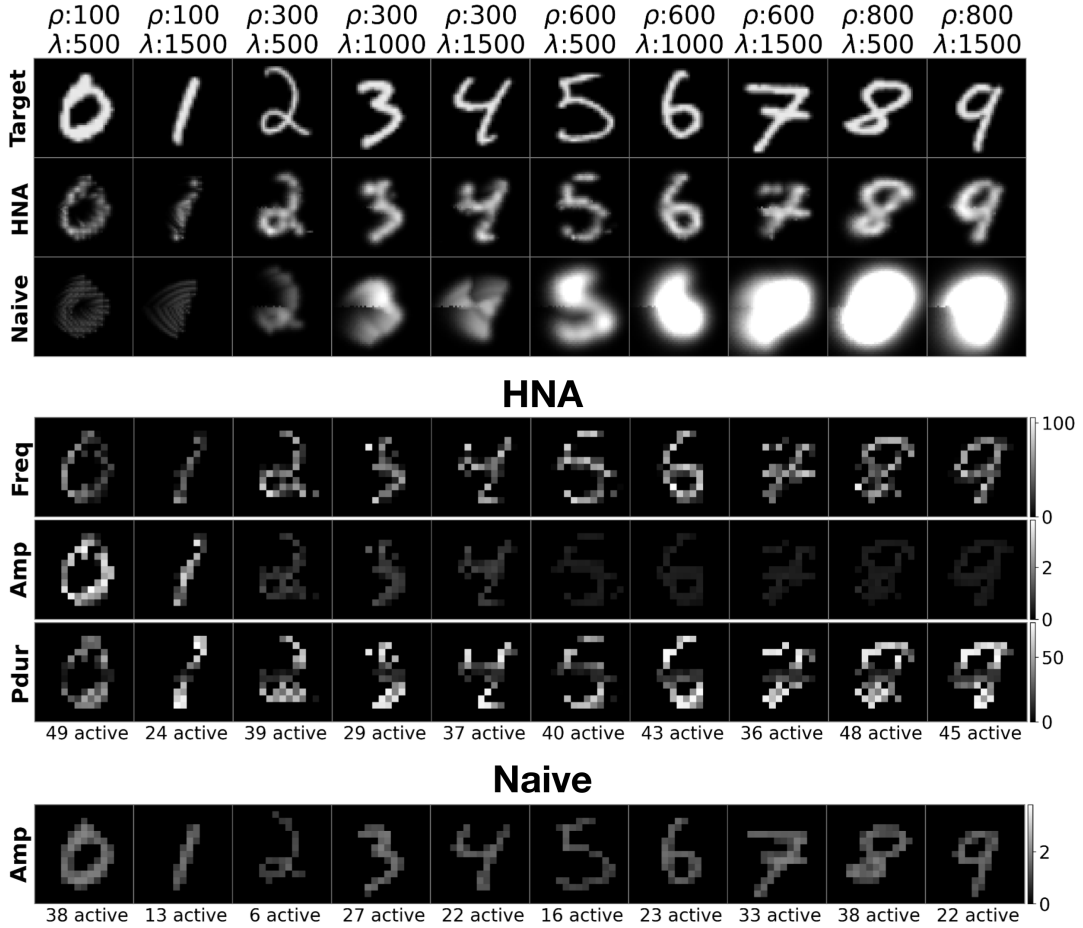


Figure A.4.1: *Top*: Example MNIST target images, and the phosphenes produced by HNA and naive encoders, encoded at various ρ and λ values. *Center*: The stimuli corresponding to the HNA phosphenes. From top to bottom, stimulus frequency (Hz), amplitude (xTh), and pulse duration (ms) are shown. The number of 'active' electrodes stimulated above threshold levels is given below each stimuli. *Bottom*: Stimuli corresponding to the naive phosphenes.

A.5 COCO Patient-to-Patient Variations

We repeated the analysis presented in Section 4.6.3 for the COCO dataset. Figure A.5.1A shows two example COCO images, encoded by both HNA and the naive encoder, across varying ρ and λ values. The heatmaps in Figure A.5.1B show the log of the joint perceptual loss across simulated patients, for both the naive and HNA encoders. To measure phosphene consistency, we performed T-SNE clustering on a subset of the COCO

images which have only 1 object. Unfortunately, T-SNE clustering of the ground-truth COCO images did not form groups corresponding to the object types (Figure A.5.1C), suggesting that the representation of object instances vary drastically across COCO images. Therefore, it was not meaningful to repeat the analysis presented in Fig. 4.4C.

Similar to the MNIST results presented in Section 4.6.3, HNA produced higher-quality representations than the naive encoder, resulting in a lower joint loss for every simulated patient. HNA performed consistently well across all simulated patients (Figure A.5.1B), with a small increase in loss for small ρ (< 100). Similar to MNIST, the naive encoder only performs well for patients with a mid-to-low ρ (≈ 200) and low λ .

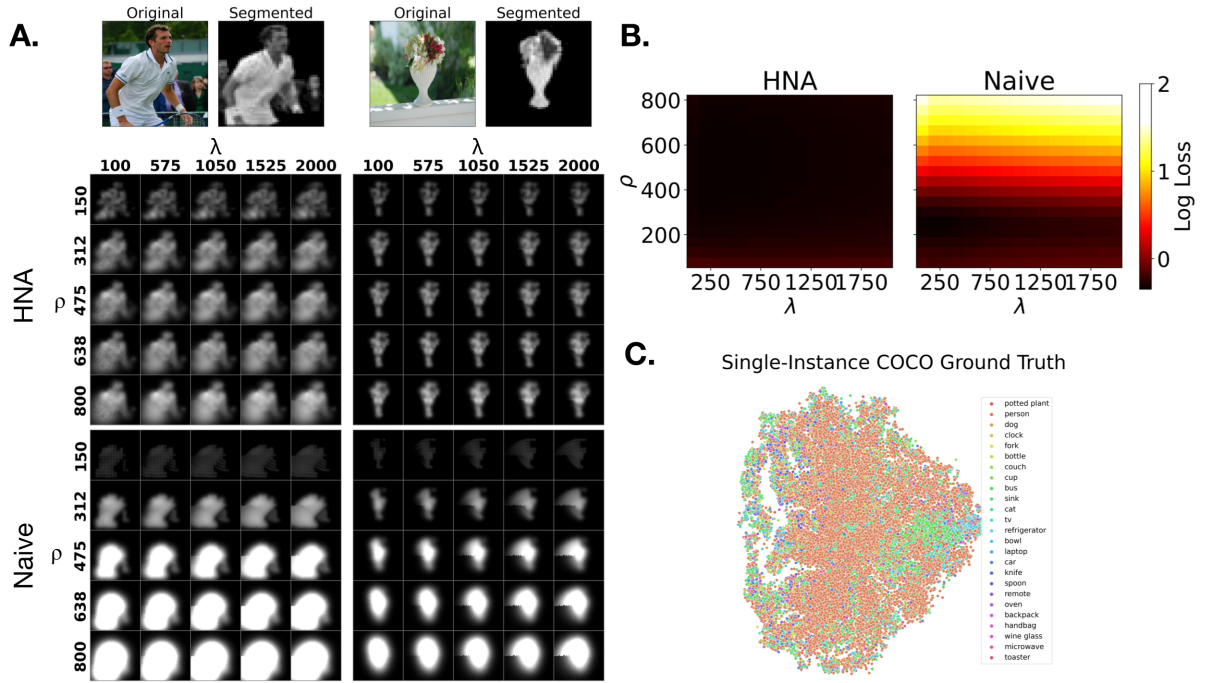


Figure A.5.1: COCO Encoder performance across simulated patients (varying ρ and λ). **A:** Phosphenes produced by HNA and Naive encoders of two example images. **B:** Heatmaps showing the log joint loss across ρ and λ for HNA and naive encoders. **C:** Ground-truth COCO images cannot be clustered using T-SNE into groups corresponding to the object types. The clustering was performed on COCO images that only contained one object.

A.6 Modeling Other Patient-to-Patient Variations

Previously, results were presented across patient-specific parameters ρ and λ , because these have the greatest impact on phosphene appearance. However, the forward model has a number of other patient-specific parameters, which HNA is also able to adapt to. For full details on all parameters of the forward model, see [22]. Out of the remaining parameters, $a2$, $a3$, and $a5$ are the most impactful on phosphene appearance. $a2$ and $a3$ modulate how much the brightness contribution from each electrode scales with increasing amplitude and frequency, respectively. $a5$ locally scales the global radial current spread ρ based on each electrodes amplitude. Figure A.6.1 (*left*) illustrates the effect of these parameters on phosphene appearance.

Figure A.6.1 compares HNA to naive encoder performance across $a2$, $a3$, and $a5$. The ranges for these parameters are based on values empirically observed in retinal prosthesis users [22]. HNA produces relatively consistent phosphenes, and outperforms the naive encoder across all conditions.

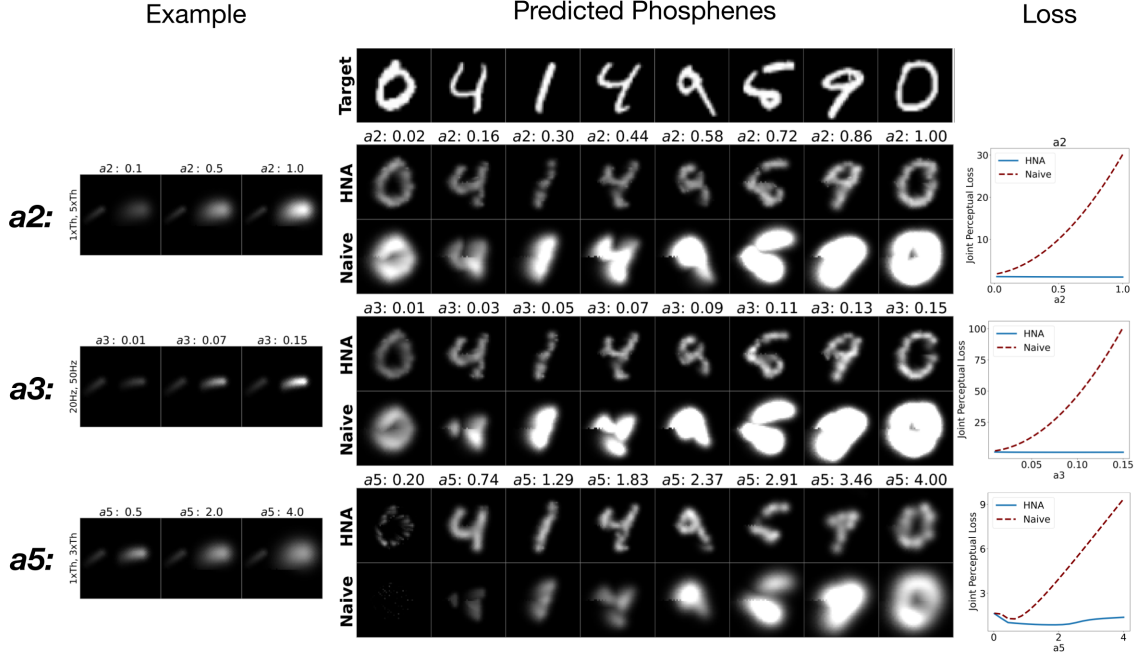


Figure A.6.1: *Left*: Examples of how a_2 , a_3 , and a_5 affect single-electrode phosphenes. a_2 modulates local brightness scaling with increasing amplitude, a_3 modulates local brightness scaling with increasing frequency, and a_5 modulates local size scaling with increasing amplitude. *Center*: Phosphenes predicted with HNA and naive encoders for varying a_2 , a_3 , and a_5 , increasing left to right. *Right*: Plot showing the joint loss across a_2 , a_3 , and a_5 for HNA (solid) and naive encoder (dashed line).

A.7 Simulating Higher-Resolution Implants

On the COCO task, HNA significantly outperformed the naive encoder, but was still unable to capture all of the detail in the images. Two of the main reasons for this are the limited spatial resolution of the implant and the patient-specific distortions from the forward model. Here, we present results from HNAs trained on implants of higher resolution, at small ρ and λ . The chosen implants are illustrated in Figure A.7.1A. For a fair comparison, each HNA was trained for only 50 epochs.

Phosphenes resulting from the HNA trained on the different implants are shown in Figure A.7.1C, and the losses across implants is plotted in Figure A.7.1B. As implant resolution increases, the phosphenes look increasingly similar to the ground truth, and

small details (e.g. facial details, textures) start to emerge.

Thus, HNAs initial failure to capture high-frequency details in the image appears to be an application-specific limitation for visual prostheses more so than a limitation of the HNA framework. For visual prostheses, learning to reconstruct the high-frequency features of complex images despite distortions and limited implant resolution remains an open problem.

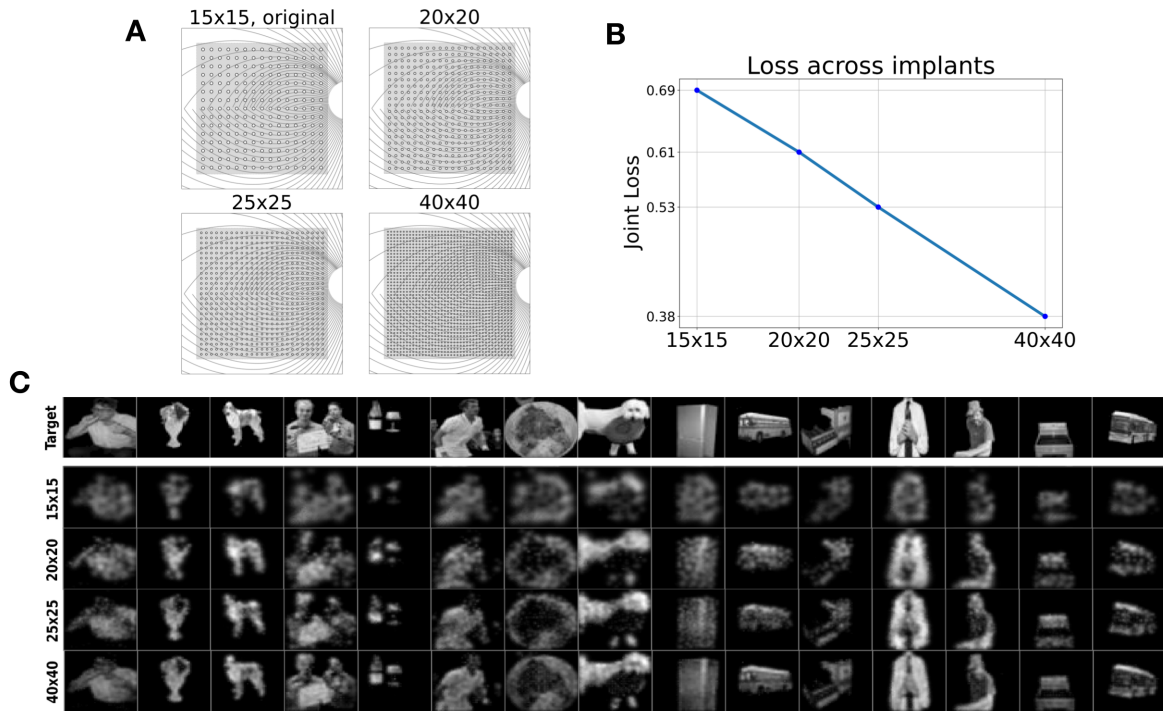


Figure A.7.1: **A**: The 4 different implants compared. The main text uses the 15×15 implant. **B**: The joint perceptual loss of HNAs trained on the different implants after 50 epochs. **C**: Example images showing the reconstructed phosphenes using each implant

A.8 Mis-Specified Patient-Specific Parameters

Due to noisy or limited patient data, there may be some uncertainty in the measured value of the patient-specific parameters ϕ . Therefore, we conducted an analysis of the consequences of incorrect patient-specific parameters on the encodings produced by HNA. Note that the true patient-specific parameters are not needed during training, so incorrect ϕ will only affect evaluation. A 'mismatch' HNA model was created, where the forward model decoder used the true patient-specific parameters ϕ , and the encoder used another set of patient-specific parameters ϕ' .

In the first experiment, ϕ' was sampled from a uniform random distribution (we again focus on only ρ and λ). The original HNA encoder, naive encoder, and mismatch HNA encoder with random ϕ' were evaluated on the MNIST test set. HNA achieved a joint loss of 0.92, the naive encoder had a joint loss of 3.13, and the mismatch HNA had a joint loss of 1.35 ± 0.003 (mean $\pm$ standard deviation across 10 random ϕ'). Thus, even if the true patient-specific parameters are completely unknown, on average randomly selecting values will still produce higher-quality encodings than the naive method.

In a second experiment, we analyzed whether there were any configurations ($\phi - \phi'$ combinations) that resulted in a worse encoding than the naive model. For the 90% of true ϕ , the mismatch model outperformed the naive model regardless of the chosen ϕ' . However, the naive model performs best at $\rho = 250$ and $\lambda = 200$. In Figure A.8.1A, we hold λ constant at 200 and, for each true ρ , plot the ranges of mis-specified ρ' for which the mismatch HNA still outperforms the naive. Figure A.8.1B shows a similar plot for varying λ , holding ρ constant at 250. Even for the naive model's ideal patients, HNA still outperforms the naive model for a large proportion of mis-specified ρ and λ .

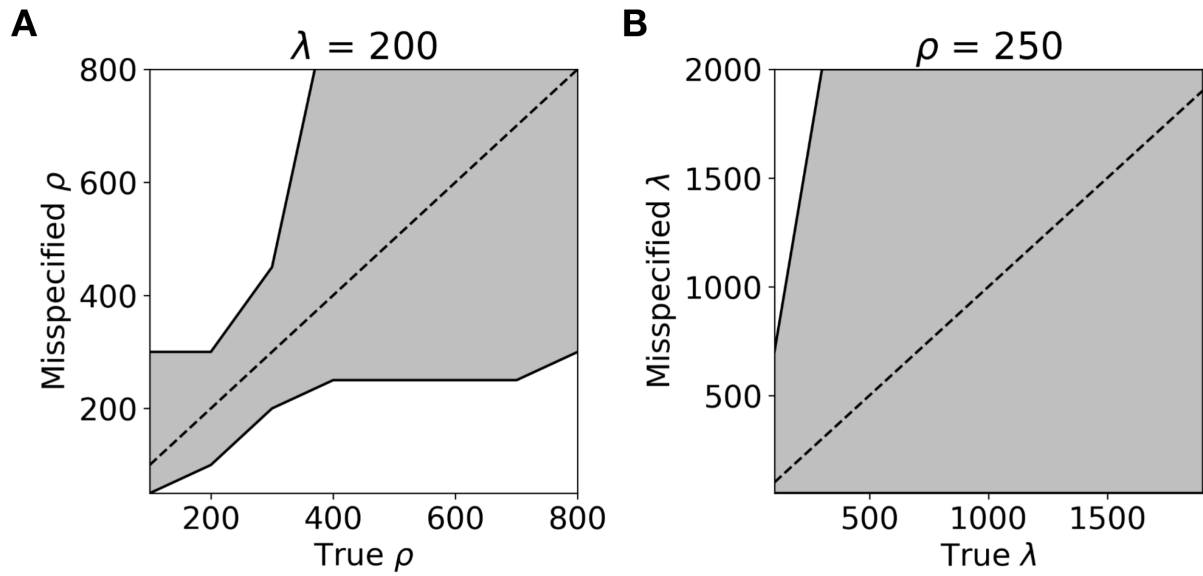


Figure A.8.1: Plots showing mis-specified HNA performance relative to the naive encoder for varying ρ (panel **A**) and λ (panel **B**). The dashed line marks the correctly specified model, and shaded area between the solid lines shows the region where the mis-specified HNA outperforms the naive encoder. Note that the naive model's ideal patient was used, with λ fixed at 200 and ρ fixed at 250, respectively.

Appendix B

Chapter 5 Appendix

B.1 Phosphene Model

B.1.1 Methods

This section describes the phosphene model used to simulate patient’s perception resulting from stimulation. The model takes in a stimulus vector $\mathbf{s} \in \mathbb{R}^{n_e \times 3}$ specifying the frequency (*freq*), amplitude (*amp*), and pulse duration (*pdur*) of a biphasic pulse train on each electrode. In addition, the model also takes in a vector of patient-specific parameters ϕ (see Table 5.1). We break these parameters down into implant parameters (x, y, rot), global parameters ($\rho, \lambda, \omega, OD_x, OD_y$), and stimulus-related parameters (a_0 - a_4); all explained below.

Exact implant locations vary patient-to-patient. The three implant parameters allow our model to account for these changes. We used a simulated implant inspired by designs of real epiretinal implants [45, 228] and those used in previous simulation studies [24]. It consists of 225 disk electrodes (radius 75 μm arranged onto a square, 15×15 grid with 400 μm spacing, initially centered over the fovea. The three implant-related parameters

translate and rotate the initial implant to be centered at (x, y) , and to be rotated by rot degrees. The implant used is depicted in Figure B.1.2, overlaid on top of a simulated map of axon nerve fiber bundles [74].

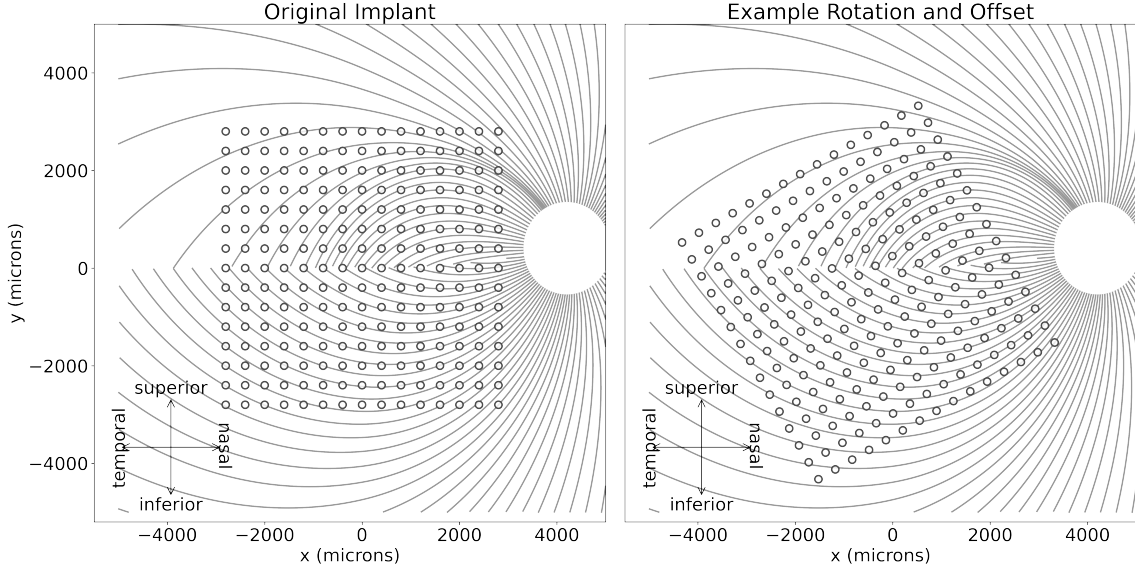


Figure B.1.1: The implant used for optimization, and an example implant after rotation and translation based on patient-specific parameters ϕ . The white circle on the right is the optic disc. Arced lines depict simulated axon nerve fiber bundles.

The remaining model parameters are inspired by various psychophysical and electrophysiological studies [17, 65, 76, 77, 123], and are summarized in the following list:

- ρ : Average phosphene size. This will be modified locally based on stimulus parameters.
- λ : Average phosphene eccentricity (a measure of phosphene elongation; not to be confused with retinal eccentricity). This will be modified locally based on stimulus parameters.
- ω : Orientation scaling factor. The orientation of phosphenes will be the orientation of the underlying axon bundle, scaled by ω (Eq. B.4).
- OD_x, OD_y : The x and y location of the patient's optic disc, into which axon nerve fiber bundles terminate.

- a_0 - a_2 : Coefficients to modulate phosphene brightness with stimulus parameters (Eq. B.1).
- a_3 : Coefficient to modulate phosphene size with stimulus parameters (Eq. B.2).
- a_4 : Coefficient to modulate phosphene eccentricity with stimulus parameters (Eq. B.3).

Each electrode's location on the retina can be determined from the implant parameters. The corresponding location in the visual field (μ_e) is determined using the retinotopic map described in Watson *et al.* [87]. Each electrode's phosphene orientation is then $\theta_e = \omega \theta_{axon}$, where θ_{axon} is the orientation of the axon nerve fiber bundle (NFB) underlying the cell (pixel). Axon NFBs are modeled as spirals originating at the optic disc and terminating at each simulated cell. These spirals follow a simulated axon map [125] based on tracings of axon trajectories in 55 human eyes. In summary, phosphene size, eccentricity, brightness, and orientation are modulated based on stimulus parameters and implant location according to the following equations:

$$b_e = a_0(amp_e)^{a_1} + a_2(freq_e) \quad (B.1)$$

$$\rho_e = \rho * a_3 * amp_e \quad (B.2)$$

$$\lambda_e = \lambda \left(\frac{pdur}{0.45} \right)^{a_4} \quad (B.3)$$

$$\theta_e = \omega * \theta_{axon} \quad (B.4)$$

The phosphene for each electrode is a multivariate Gaussian blob, centered at the electrodes location in visual field (μ_e), and with covariance matrix Σ_e constructed such that the resulting phosphene will have brightness b_e , size ρ_e , eccentricity λ_e , and orientation θ_e , as shown in the following equations (repeated from main text for convenience):

$$b(x, y) = 2\pi b_e \det(\Sigma_e) \mathcal{N}([x, y]^\top | \mu_e, \Sigma_e), \quad (\text{B.5})$$

The covariance matrix $\Sigma_e = \mathbf{R}\Sigma_0\mathbf{R}^T$ is calculated from the eigenvalue matrix Σ_0 and a rotation matrix $\mathbf{R}$:

$$\Sigma_0 = \begin{bmatrix} \alpha_e^2 & 0 \\ 0 & \beta_e^2 \end{bmatrix}, \quad R = \begin{bmatrix} \cos \theta_e & -\sin \theta_e \\ \sin \theta_e & \cos \theta_e \end{bmatrix}.$$

The eigenvalues α_e and β_e depend on the intended phosphene area (ρ_e) elongation (λ_e), and a constant ϵ (set to e^{-2}):

$$\alpha_e^2 = -\frac{\rho_e \sqrt{1 - \lambda_e^2}}{2\pi \ln \epsilon}, \quad \beta_e^2 = -\frac{\rho_e}{2\pi \ln \epsilon \sqrt{1 - \lambda_e^2}}.$$

This formulation guarantees that the Gaussian blob, when thresholded using ϵ , will have the intended area, orientation, eccentricity, and brightness.

Blobs from individual electrodes are summed into a global percept. This linear summation is supported by recent studies, which have shown that percepts from multi-electrode stimulation are often linearly related to the percepts from stimulation on the individual electrodes [229]. Although the sum across electrodes is linear, modulating the size and eccentricity of phosphenes makes the final result a nonlinear function of stimulus parameters, preventing analytic inversion.

B.1.2 Evaluation

Our model is motivated by similar anatomical and psychophysical phenomena as the previous state-of-the-art model for epiretinal prostheses [22], but its formulation allows for favorable computational properties. In comparison, our model is on average 45x faster

to run, and consumes about 120x less GPU memory. These computational benefits are the main reason a new model was necessary, and enables a more advanced deep stimulus encoder by allowing training with larger encoder models, longer training duration, and larger batch sizes.

Nonetheless, we also verified that the model produces state-of-the-art predictions, as described in Section 5.5.1. Despite its similar design, our model achieves much better scores on the Beyeler *et al.* [17] evaluation for phosphene shape. This is likely because our formulation allows much tighter control of phosphene shape attributes (e.g., size, eccentricity), allowing (for the first time) positive R^2 on shape descriptors for held-out electrodes. Our model performs similarly to the previous state-of-the-art model on the Granley *et al.* [22] evaluation, which is to be expected given that the equations modulating phosphene appearance with stimulus parameters in both models are very similar. Fig. B.1.2 reproduces the plots from Figures 4A-C and 5 of [22], but with our proposed model included. These figures show the brightness or size rating from Argus II patient(s) as stimulus parameters vary [77, 127], in subjective units. For brightness, ‘10’ means the same as the reference pulse, ‘20’ means twice as bright, etc. For size, ‘1’ means the same as reference pulse, ‘2’ means twice as large (notation matching [22, 77, 127]).

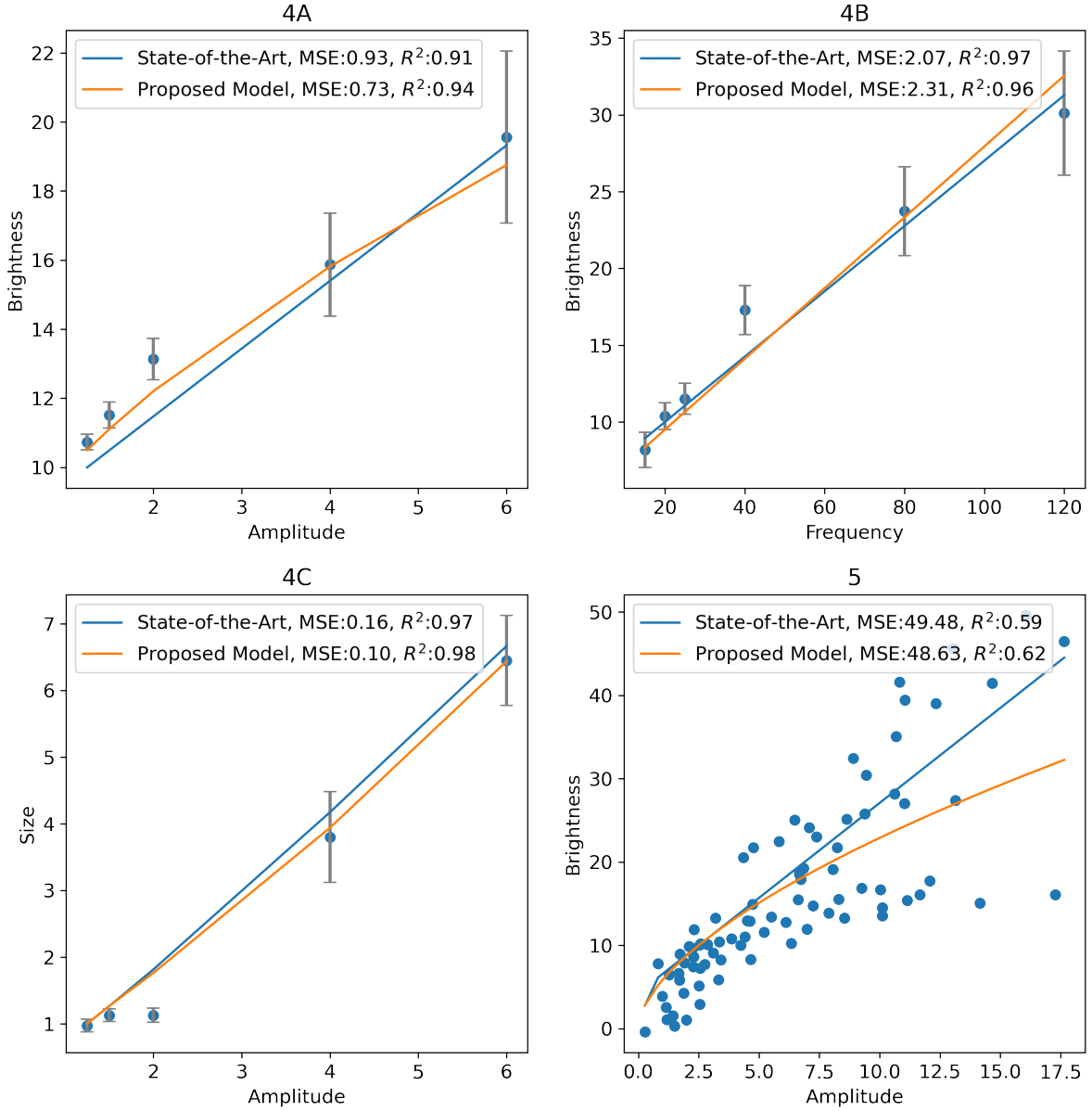


Figure B.1.2: Evaluation of phosphene brightness and size as stimulus parameters vary. Reproduced from [22], but with our proposed phosphene model included. State-of-the-art denotes the phosphene model from [22]. Units are subjective, in comparison to a reference pulse (i.e. brightness of 20 means twice as bright, size of 3 means 3 times as bright) [77]

B.2 Deep Stimulus Inversion

B.2.1 Architecture

The encoder architecture described in Section 5.4 is illustrated in Fig. B.2.1.

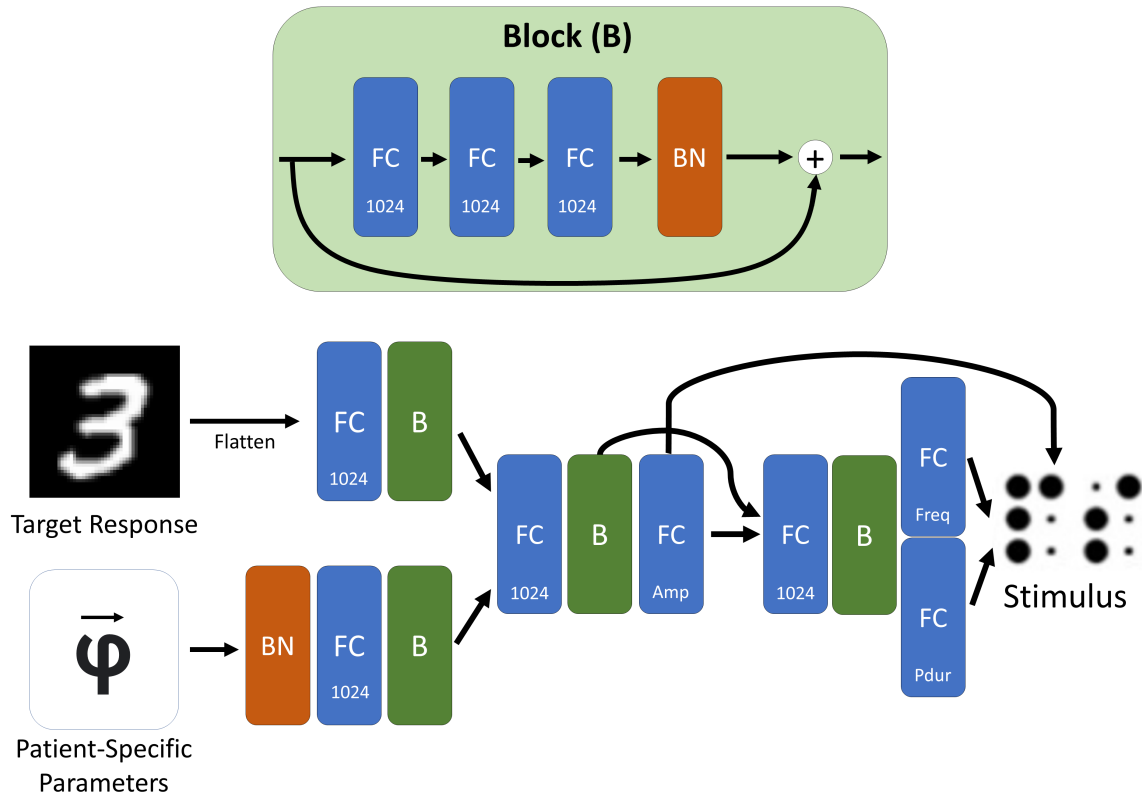


Figure B.2.1: Deep stimulus encoder architecture. FC: fully connected layer, BN: batch normalization layer, B: block of layers (shown at top). Two arrows merging into one fully connected layer denotes concatenation. 45M total parameters.

B.3 Human-in-the-Loop Optimization

B.3.1 Kernel Selection and Hyperparameters

This section gives more details on fitting the hyperparameters for the Gaussian process (GP) kernel used in preferential Bayesian optimization (PBO). As stated previously, the performance of PBO crucially depends on the GP kernel and its hyperparameters, which encode our prior assumptions about the latent preference function. To select hyperparameters, we used a transfer learning strategy, selecting hyperparameters that generalized best within a small validation group of simulated patients (see Appendix B.3.2 for discussion).

We simulated 600 random duels (ϕ_1 and ϕ_2 chosen randomly) on each of 10 simulated patients. For each patient, we fit four commonly used kernels (Squared Exponential, Squared Exponential with Automatic Relevance Determination (ARD), Matérn 3/2, and Matérn 5/2) and inferred hyperparameters using type II maximum likelihood estimation [230]. The bounds for each hyperparameter were $[\exp(-10), \exp(10)]$. For each of these candidate kernel-hyperparameter pairs, we fit a GP with the corresponding kernel and hyperparameters to 50 training duels for each of the other 9 patients. Then, the performance of the candidate GP was evaluated on the remaining 550 data points using Brier score [**brier1950verification**], a commonly used metric measuring the accuracy of probabilistic predictions:

$$BS = \frac{1}{n} \sum_{i=1}^n (y_{true} - y_{pred})^2, \quad (\text{B.6})$$

where y_{true} is the true duel outcome (1 or 0) as decided by the simulated patient, and y_{pred} is the probability of ϕ_1 being selected as the winner (corresponding to outcome of 1), as predicted by the Gaussian process.

The kernel and hyperparameters with the lowest Brier score, averaged across all 9 other patients, were selected (Matérn 5/2). To verify that this kernel performed well, we also ran human-in-the-loop optimization (HILO) for 20 random simulated patients, using the best hyperparameters for each of the four previously mentioned kernels. The results are shown in Figure B.3.1. The Matérn 5/2 kernel performed slightly better than the Matérn 3/2 kernel, and significantly better than the ARD kernel. While performance was similar to the Squared Exponential kernel, we ultimately selected Matérn 5/2 due to its lower Brier score.

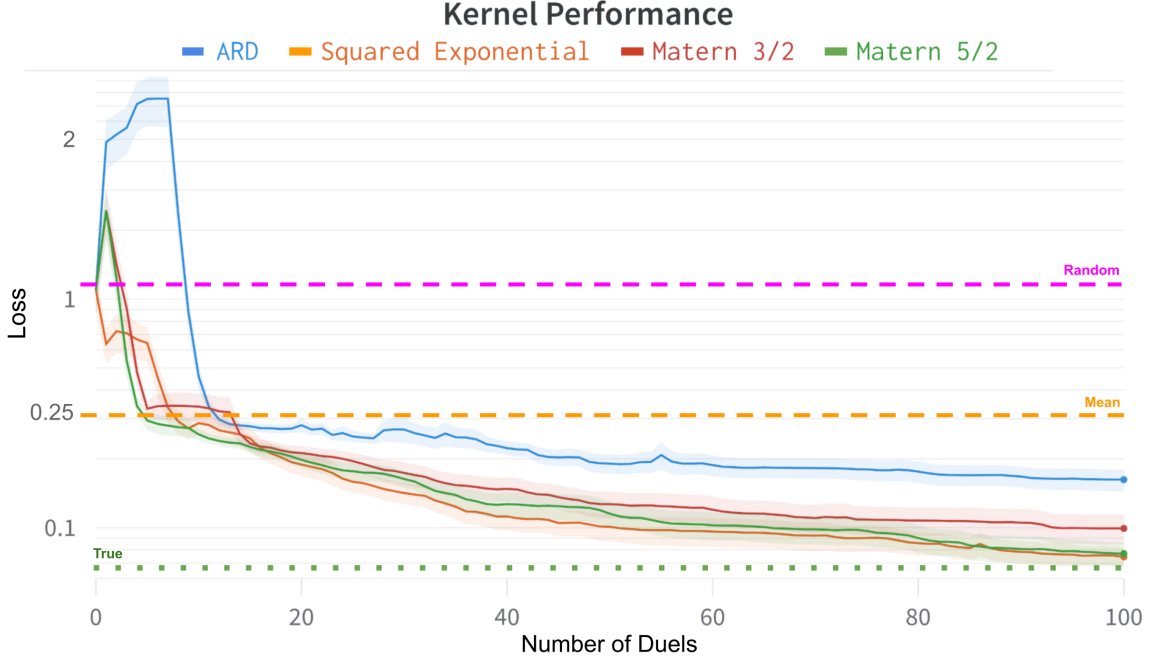


Figure B.3.1: Joint perceptual loss (y axis, log scale) plotted throughout HILO with different Gaussian process kernels. Error bars denote SEM.

B.3.2 Hyperparameter Optimality

The Gaussian process kernel hyperparameters selected with our transfer learning strategy performed well in our simulations, leading to higher-quality patient-specific

stimulus encodings (Figure 5.4). This transfer learning approach was chosen to match a clinical setting, where limited human data availability. However, it is certainly possible that better performing hyperparameters exist. To investigate the optimality of our hyperparameters, we examine two other strategies: an ideal case where ‘patient-optimal’ hyperparameters are used for each patient, and an online strategy where hyperparameters are updated during optimization with each patient,

Patient-optimal Hyperparameter Selection In this strategy, hyperparameters were precomputed for each patient by simulating 200 random duels. Kernel parameters were again fit using type II maximum likelihood estimation [230]. These ‘patient-optimal’ hyperparameters were then used during HILO for the patient.

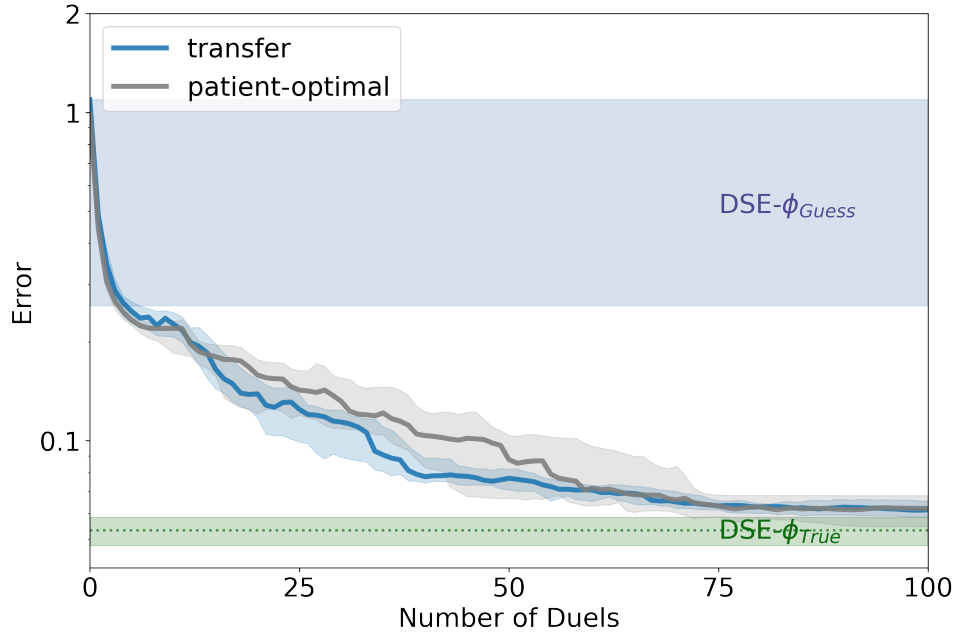


Figure B.3.2: Reconstruction error using Gaussian process kernel hyperparameters selected via transfer learning and ‘patient-optimal’ strategies throughout HILO.

Optimization results for 20 random patients are shown in Figure B.3.2. Both the transfer and the patient-optimal settings led to similar performance, and there was no significant difference between the final reconstruction errors ($p > .05$, two-sided paired

t-test). Thus it seems the transfer learning strategy indeed selected hyperparameters that generalize well for new patients. Note that in general the transfer learning selection strategy is more practically applicable than the patient-optimal strategy since it does not require a long calibration period, but the patient-optimal strategy is an alternative that could be used for the first human subjects when no human data (only simulated data) is available.

Online hyperparameters selection While it is common to keep kernel hyperparameters constant during optimization [185, 231, 232], online optimization is an alternative, where kernel parameters are periodically re-fit to the patient data during optimization. We tested online optimization where hyperparameters were recalculated with an update period of 1, 5, 10, or 20 duels, or were never updated. In all settings, the initial hyperparameters were chosen using the transfer learning strategy.

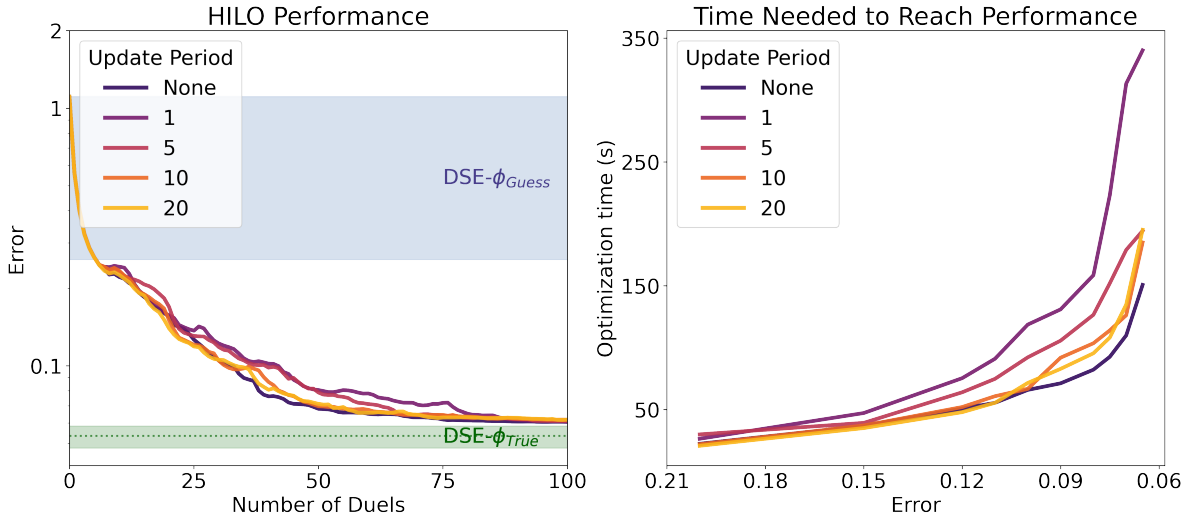


Figure B.3.3: Reconstruction error (left) and Bayesian optimization time required to reach a specified reconstruction error (right) for HILO with online updates of GP hyperparameters.

The results for online hyperparameters optimization are shown in Figure B.3.3. All update periods eventually converged to similar performance as with the transfer learning

strategy (*i.e.* never updating hyperparameters), but since recalculating hyperparameters is costly, online hyperparameter updates increased the optimization time required to reach a desired performance (Figure B.3.3, *right*). This suggests that just using the transfer learning hyperparameters is a better strategy than online updates.

B.3.3 Acquisition function

The acquisition function is responsible for choosing ϕ_1 and ϕ_2 for each duel, and must balance exploration of the search space with exploiting values of ϕ that are expected to work well. The Maximally Uncertain Challenge (MUC) [191] acquisition presented in the main text was initially compared against 2 other top-ranking [191] acquisition functions: Bivariate Expected Improvement (Bivariate EI) [233], and Dueling Upper Credibility Bound (Dueling UCB) [234]. In addition, we also compare against a baseline acquisition, where ϕ_1 and ϕ_2 are chosen randomly for each duel. We ran HILO for the same 20 random patients for each acquisition.

The joint perceptual loss throughout optimization for each acquisition is presented in Figure B.3.4. All of the tested acquisition functions dramatically outperformed the random baseline, which converged to a value near the mean DSE without HILO. Although MUC and Dueling UCB performed similarly, we ultimately selected MUC due to its slightly lower final loss.

B.3.4 Batch Bayesian Optimization

While the proposed Bayesian optimization strategy of sequentially presenting the user with stimuli would likely not be prohibitively time consuming (about 17 minutes with 100 duels, 10 seconds per duel), it is possible that batch Bayesian optimization could speed up optimization. We consider two formulations of batch Bayesian optimizations: 1) the

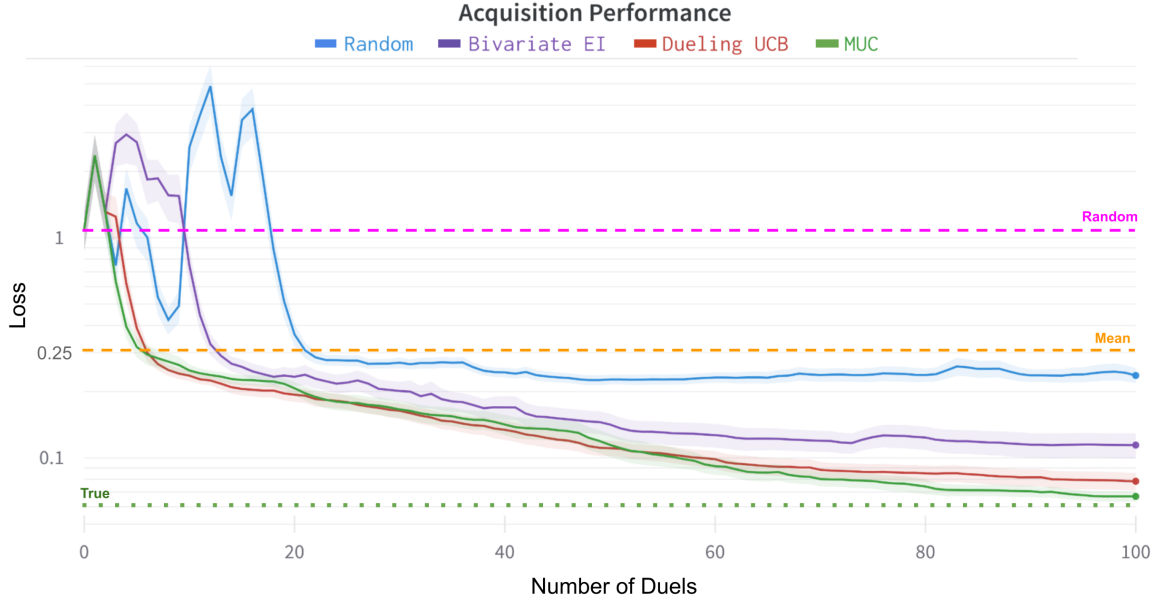


Figure B.3.4: Joint perceptual loss (y axis, log scale) plotted throughout HILO with different acquisition functions. Error bars denote SEM.

user selects their preference from a batch of stimuli in each iteration, and 2) each duel still has only two options, but batches of duels are precomputed to save on updated the Gaussian process posterior and the acquisition function time between duels.

Option 1 is theoretically more ideal, since more information acquired in each comparison would hopefully allow for fewer comparisons. However, this would require the patient to remember an entire batch of stimuli before making a comparison, and in practice phosphenes can be very difficult to discriminate [59]. It is difficult to evaluate the effect of this cognitive burden on simulated patients, and we thus leave it to future work to consider whether the parallelization of data acquisition would make up for the increased difficulty of the task.

Therefore, option 2 was tested by precomputing batches of 1 (*i.e. original*), 3, 6, or 10 duels at a time. We used a batch variant of the maximally-uncertain challenge acquisition function [191]. Results for 20 simulated patients are shown in Figure B.3.5. Precomputing

batches reduced the optimization time required to reach a desired performance, but at the cost of requiring more duels. In practice, the optimal batch size could be determined by balancing the time required per duel with the time required for optimization. Faster acquisitions (*e.g.* KernelSelfSparring [sui2017multi]) could further reduce optimization time.

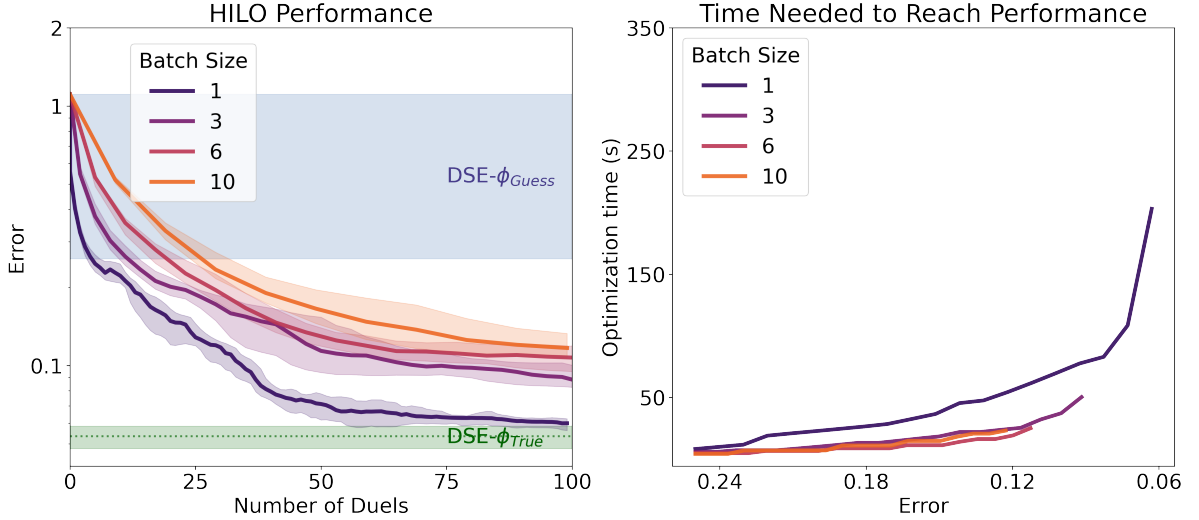


Figure B.3.5: Reconstruction error (left) and Bayesian optimization time required to reach a specified reconstruction error (right) for HILO with different batch sizes.

B.3.5 Baselines and Robustness

In this section, we provide further details for the naive encoder we compared against, and the robustness experiments.

Naive Encoder The naive encoder is the encoding strategy currently used in commercial epiretinal prostheses [45]. This encoder operates under the assumption that each electrode can be thought of as a pixel in an image. The optimal stimulus under this assumption is therefore simply a downsampled version of the target image. The frequency and pulse duration are constant across all electrodes. This naive encoder has been previously shown

to be suboptimal [24], but we still include it as a comparison to the currently used encoding strategy.

Robustness In section 5.5.4 we evaluate the robustness of human-in-the-loop optimization (HILO) to misspecifications in the forward model. Here, we provide specific details on the implementation of these misspecifications, and how we adapted the baseline encoders to the misspecified scenarios.

- **Axon Trajectories:** The simulated axon map from [125] has two parameters, β_{inf} and β_{sup} , which control the axon trajectories in the inferior and superior retina, respectively. [125] also reports the observed ranges for these parameters: $\beta_{sup} \in [-2.5, -1.3]$ and $\beta_{inf} \in [0.1, 1.3]$. The unmodified model uses the centers of these ranges. Under misspecification, we randomly set both β_{sup} and β_{inf} to one of these bounds for each patient.
- **Thresholds:** Threshold is the amplitude at which a phosphene becomes visible to a patient 50 % of the time. In epiretinal prostheses, thresholds are notoriously noisy, and vary significantly across electrodes, patients, and over time [194]. While some progress has been made towards predicting these thresholds [235], most state-of-the-art models assume that thresholds are known.

With this misspecification, the assumed threshold on each electrode was modified by a random but systematic amount. Specifically, the threshold on each electrode was randomly selected to be between $\frac{1}{2}x$ and $2x$ its original value for the 100% condition and between $\frac{1}{4}x$ and $4x$ its original value for the 300% condition.

- **Out of Distribution:** It is also possible that a new patient does not fall within our assumed ranges. Thus, we tested a variant where the true patient-specific parameters ϕ were sampled from outside the ranges in Table 5.1. Specifically, each

parameter was sampled to be 0-50% above or below the specified range (some parameters were clipped to stay within defined ranges, *e.g.*, λ cannot be outside of $[0, 1)$).

During HILO, the acquisition functions have specified bounds that constrain candidate ϕ . We therefore tested two variants, one where PBO was allowed to expand the bounds, and another where it was confined to within its original bounds. The end results were similar in terms of DSE performance, so the variant with its original bounds is presented in the main text.

DSE- ϕ_{Guess} is our best approximation of what a DSE would guessed patient-specific parameters would perform, and is bounded by the performance of a DSE with mean ϕ , and random ϕ from the ranges in Table 5.1. For each of the misspecifications, the mean and random ϕ baseline DSEs are still encoded with the same ϕ as in the unchanged patient, but since the phosphene model for the patient is changed, the resulting loss is different. DSE- ϕ_{True} is still parameterized with the patients true ϕ , however, the true ϕ are no longer a perfect description of the misspecified patient. This is shown by the fact that HILO surpasses the true encoders performance: under the misspecified model, there exists some other ϕ which leads to percepts with improved perceptual quality when decoded using the misspecified phosphene model, compared to those encoded with the true ϕ . This highlights the robustness of optimization based on user preferences.

Appendix C

Chapter 6 Appendix

C.1 Supplementary Figures

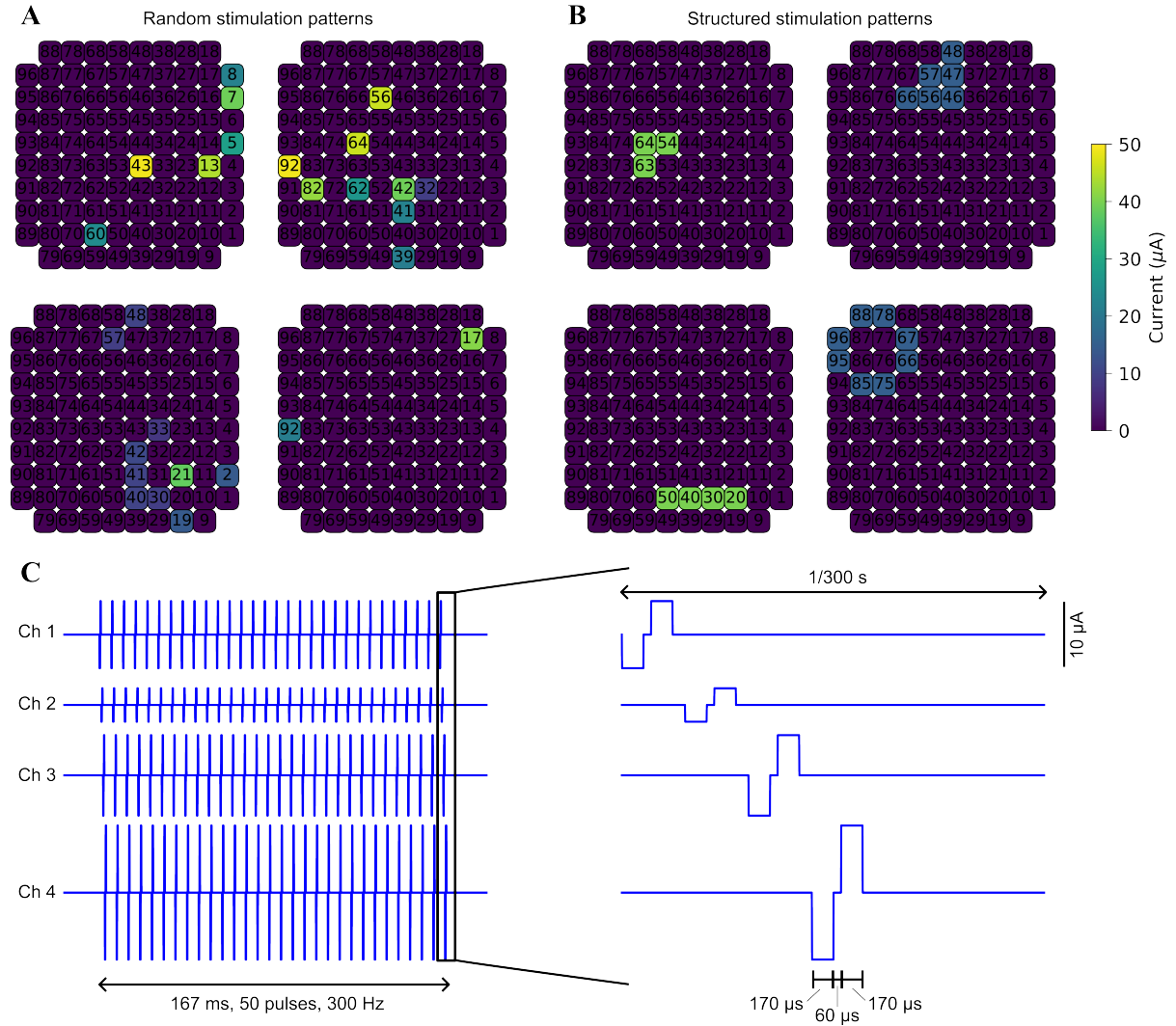


Figure 1: **Stimulation patterns.** A) Examples of random stimulation patterns. B) Examples of structured stimulation patterns. C) Example of a stimulation train from 4 electrodes (left) and zoom into the stimulation pulses (right).

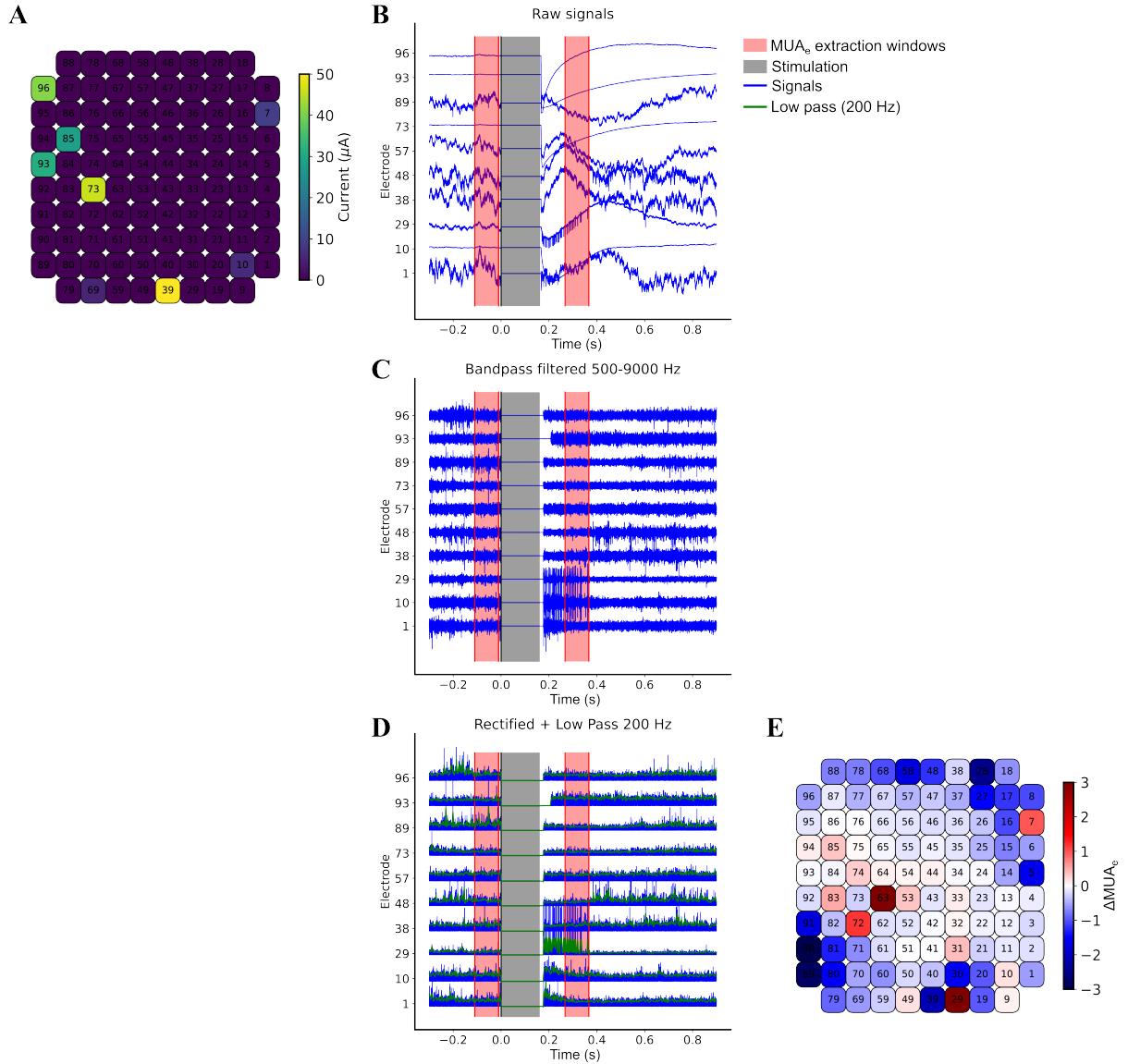


Figure 2: **Extraction of the neural activity.** A) Stimulation pattern. B) Raw signal. The black line represents the start of the stimulation, red lines represent the window used to extract the activity. C) Bandpass filtered signal (500-9000 Hz) D) Rectified signal (blue traces) and low pass filtered signal (200 Hz, green traces). E) Neural activity (ΔMUA_e).

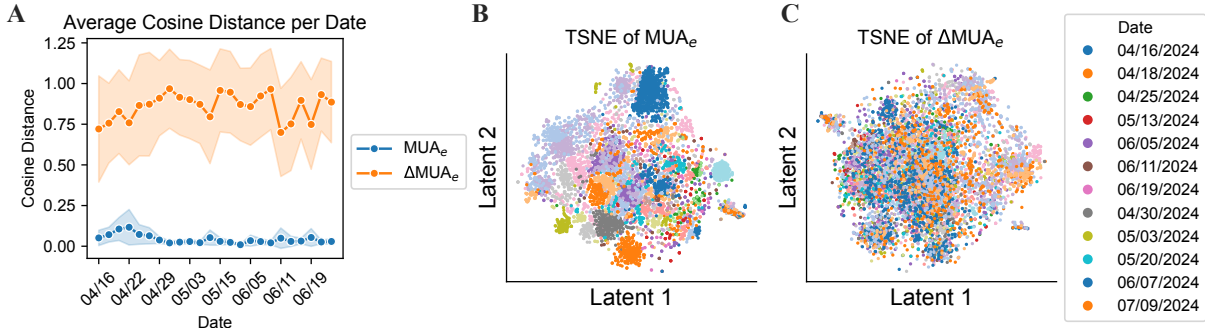


Figure 3: ΔMUA_e and post-stimulus MUA_e comparison of cosine distance (A) and t-distributed stochastic neighbor embedding (t-SNE) (B,C). We found that ΔMUA_e exhibited a higher cosine distance between points for a given day across all days, compared to the post-stimulation MUA_e (Supplementary Figure 3A). This suggests that the raw recordings contain a significant correlation structure dominated by a day-based baseline, which is mitigated in the ΔMUA_e . Visualization of the data using t-distributed stochastic neighbor embedding (t-SNE) further supported this finding. The post-stimulation MUA_e showed much stronger day-wise clustering compared to ΔMUA_e (Supplementary Figure 3B,C). This indicates that the ΔMUA_e may be less dependent on the short-term neural state of a given day.

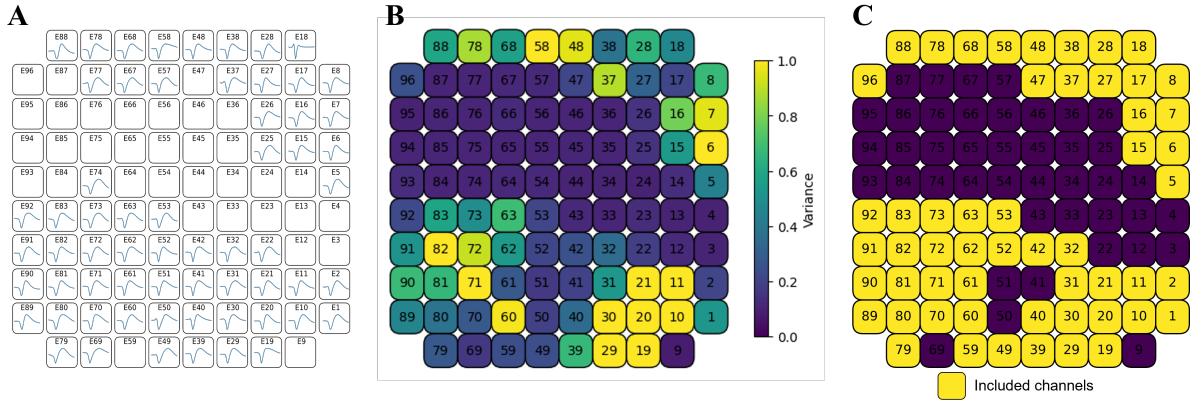


Figure 4: A) Spiking electrodes with manual spike sorting. B) Variance of the neural response to stimulation for each electrode. C) Included electrodes.

NATURAL TARGET RESPONSES

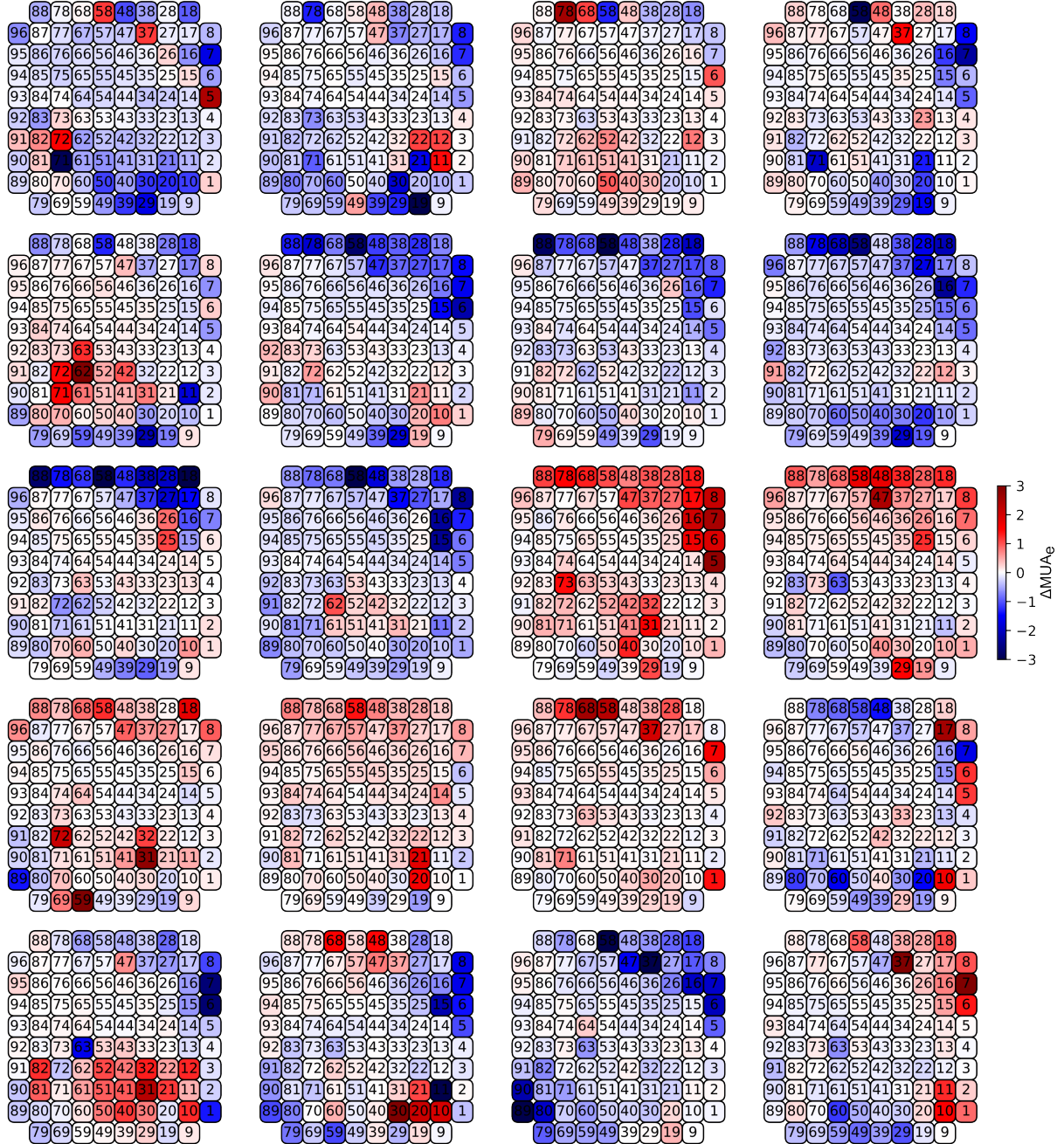


Figure 5: Natural Target Responses.

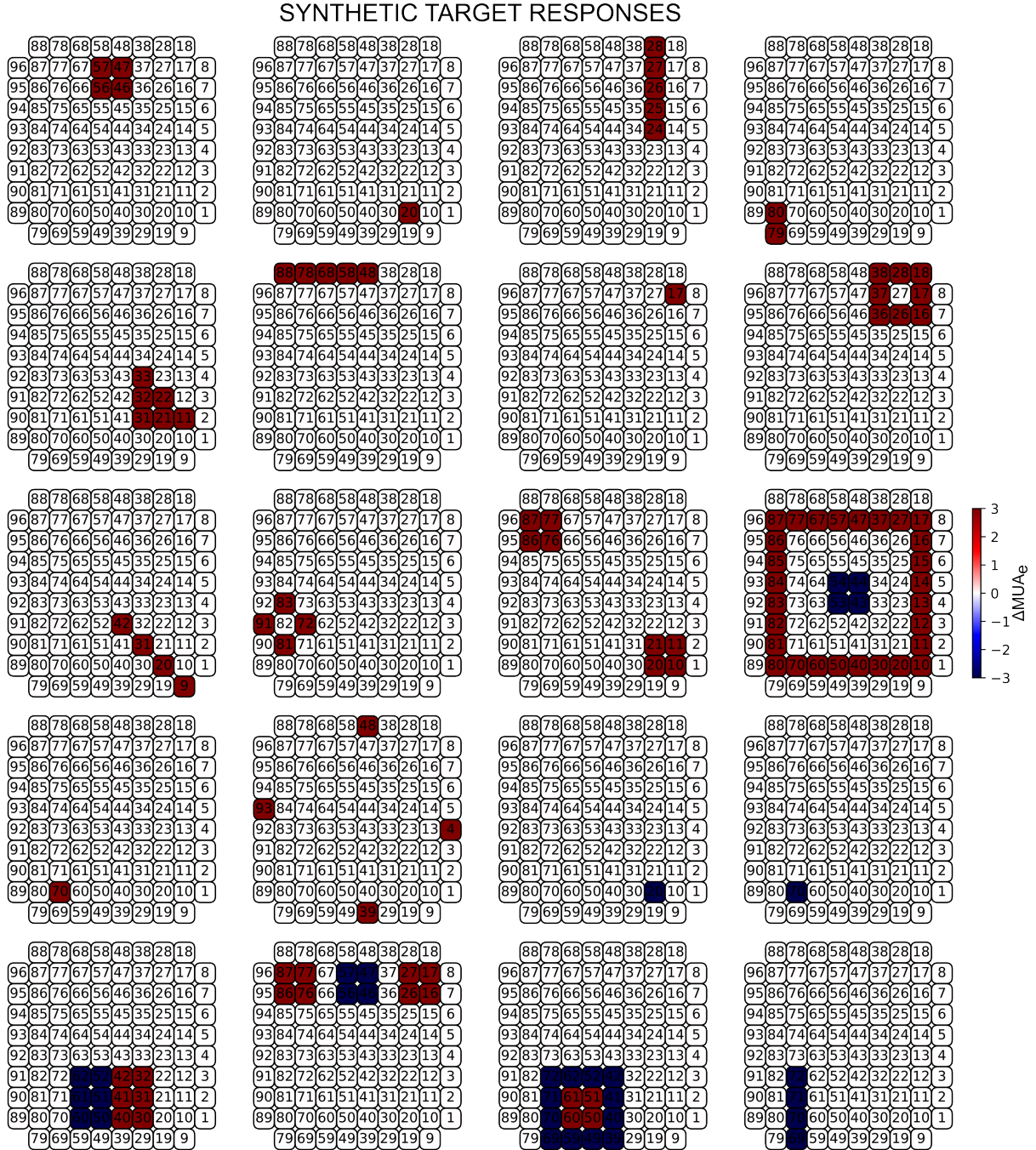


Figure 6: Synthetic Target Responses.

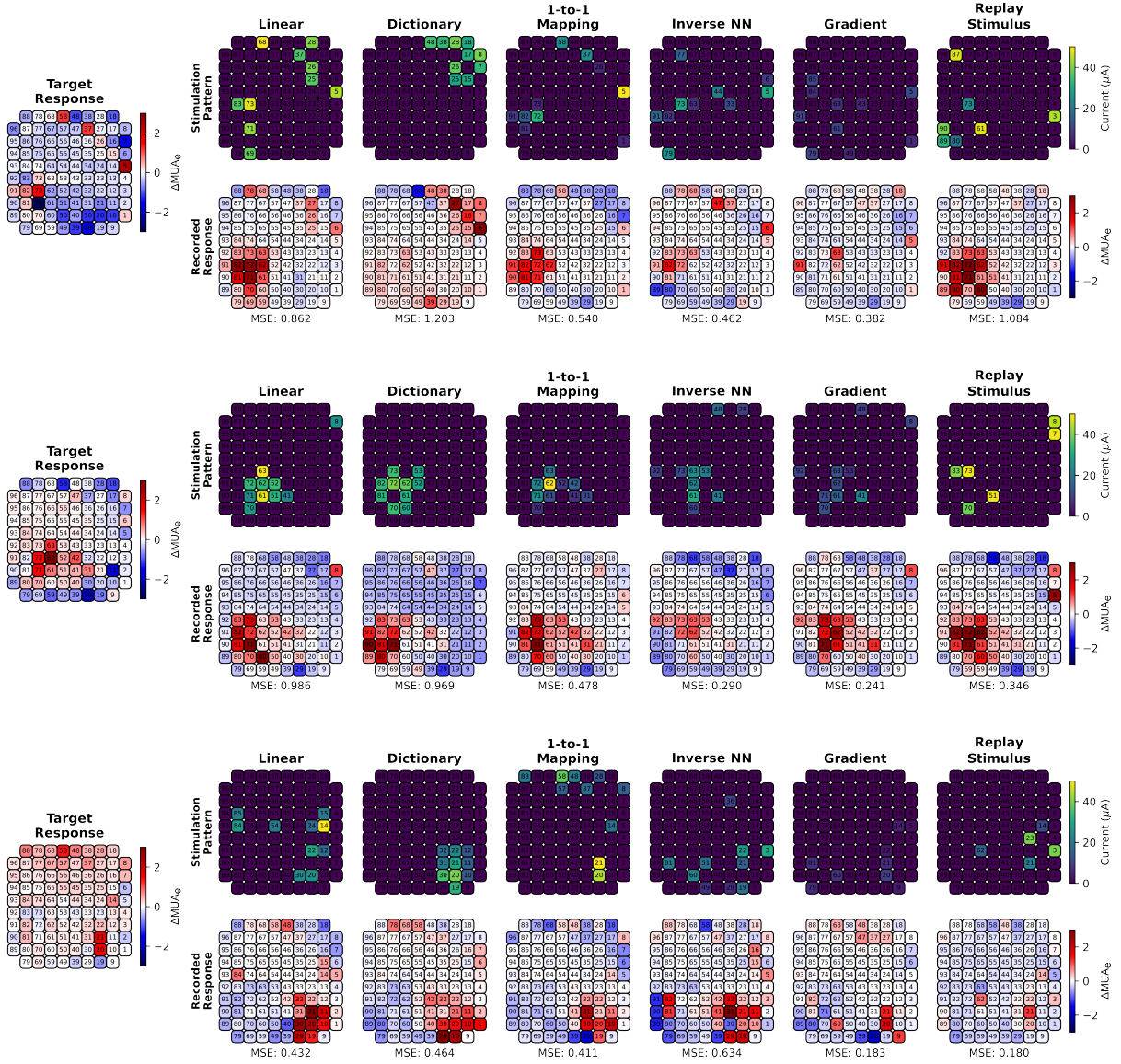


Figure 7: **Natural Target Responses examples.** For 3 examples of target activities (left): stimulation pattern proposed by each method (top) and neural response obtained (bottom).

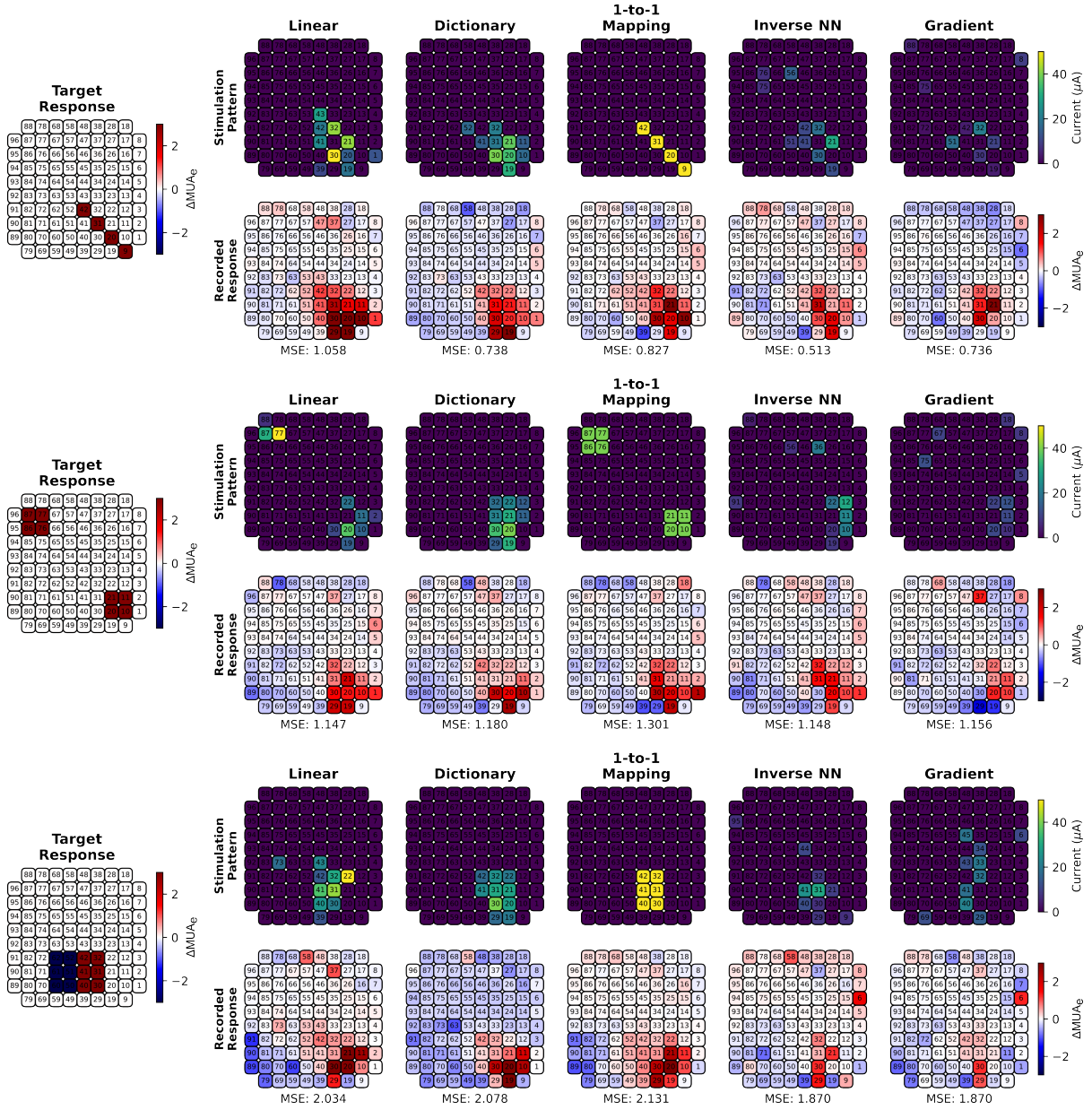


Figure 8: **Synthetic Target Responses examples.** For 3 examples of target activities (left): stimulation pattern proposed by each method (top) and neural response obtained (bottom).

Bibliography

- [1] Philip M. Lewis et al. “Restoration of vision in blind individuals using bionic devices: A review with a focus on cortical visual prostheses”. In: *Brain Research* 1595 (Jan. 2015), pp. 51–73. ISSN: 0006-8993. DOI: 10.1016/j.brainres.2014.11.020. URL: <https://www.sciencedirect.com/science/article/pii/S0006899314015674> (visited on 12/02/2024).
- [2] Xi Liu et al. “A narrative review of cortical visual prosthesis systems: the latest progress and significance of nanotechnology for the future”. In: *Annals of Translational Medicine* 10.12 (June 2022), p. 716. ISSN: 2305-5839. DOI: 10.21037/atm-22-2858. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9279795/> (visited on 12/02/2024).
- [3] Léo Pio-Lopez, Romanos Poulkouras, and Damien Depannemaecker. “Visual cortical prosthesis: an electrical perspective”. In: *Journal of Medical Engineering & Technology* 45.5 (July 2021). Publisher: Taylor & Francis, pp. 394–407. ISSN: 0309-1902. DOI: 10.1080/03091902.2021.1907468. URL: <https://doi.org/10.1080/03091902.2021.1907468> (visited on 12/02/2024).
- [4] Fan-Gang Zeng et al. “Cochlear Implants: System Design, Integration, and Evaluation”. In: *IEEE Reviews in Biomedical Engineering* 1 (2008). Conference Name: IEEE Reviews in Biomedical Engineering, pp. 115–142. ISSN: 1941-1189. DOI:

- 10.1109/RBME.2008.2008250. URL: <https://ieeexplore.ieee.org/abstract/document/4664429> (visited on 12/02/2024).
- [5] Sharlene N. Flesher et al. “Intracortical microstimulation of human somatosensory cortex”. en. In: *Science Translational Medicine* 8.361 (Oct. 2016). ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.aaf8083. URL: <https://www.science.org/doi/10.1126/scitranslmed.aaf8083> (visited on 12/02/2024).
- [6] David A. Bjånes and Chet T. Moritz. “A Robust Encoding Scheme for Delivering Artificial Sensory Information via Direct Brain Stimulation”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.10 (Oct. 2019). Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering, pp. 1994–2004. ISSN: 1558-0210. DOI: 10.1109/TNSRE.2019.2936739. URL: <https://ieeexplore.ieee.org/abstract/document/8809212> (visited on 12/02/2024).
- [7] Luis Garcia-Larrea and Roland Peyron. “Motor cortex stimulation for neuropathic pain: From phenomenology to mechanisms”. eng. In: *NeuroImage* 37 Suppl 1 (2007), S71–79. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2007.05.062.
- [8] Joel S. Perlmuter and Jonathan W. Mink. “DEEP BRAIN STIMULATION”. en. In: *Annual Review of Neuroscience* 29. Volume 29, 2006 (July 2006). Publisher: Annual Reviews, pp. 229–257. ISSN: 0147-006X, 1545-4126. DOI: 10.1146/annurev.neuro.29.051605.112824. URL: <https://www.annualreviews.org/content/journals/10.1146/annurev.neuro.29.051605.112824> (visited on 12/02/2024).
- [9] Cordelia Erickson-Davis and Helma Korzybska. “What do blind people “see” with retinal prostheses? Observations and qualitative reports of epiretinal implant users”. en. In: *PLOS ONE* 16.2 (Feb. 2021). Publisher: Public Library of Science, e0229189. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0229189. URL: <https://doi.org/10.1371/journal.pone.0229189>

- [//journals.plos.org/plosone/article?id=10.1371/journal.pone.0229189](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229189)
(visited on 12/02/2024).
- [10] Craig D. Murray. “Embodiment and Prosthetics”. en. In: *Psychoprosthetics*. Ed. by Pamela Gallagher, Deirdre Desmond, and Malcolm MacLachlan. London: Springer, 2008, pp. 119–129. ISBN: 978-1-84628-980-4. DOI: 10.1007/978-1-84628-980-4_9. URL: https://doi.org/10.1007/978-1-84628-980-4_9 (visited on 05/19/2022).
- [11] Michael P Barry et al. “Video-mode percepts are smaller than sums of single-electrode phosphenes with the Orion® visual cortical prosthesis”. In: *Investigative Ophthalmology & Visual Science* 61.7 (June 2020), p. 927. ISSN: 1552-5783.
- [12] Gislin Dagnelie et al. “Constructing a phosphene map for the inaugural recipient of the intracortical visual prosthesis (ICVP)”. In: *Investigative Ophthalmology & Visual Science* 64.8 (June 2023), p. 5520. ISSN: 1552-5783.
- [13] James D. Weiland and Mark S. Humayun. “Visual Prosthesis”. In: *Proceedings of the IEEE* 96.7 (July 2008), pp. 1076–1084. ISSN: 1558-2256. DOI: 10.1109/JPROC.2008.922589. URL: <https://ieeexplore.ieee.org/document/4539488/> (visited on 09/03/2025).
- [14] Eduardo Fernandez. “Development of visual Neuroprostheses: trends and challenges”. en. In: *Bioelectronic Medicine* 4.1 (Aug. 2018), p. 12. ISSN: 2332-8886. DOI: 10.1186/s42234-018-0013-8. URL: <https://doi.org/10.1186/s42234-018-0013-8> (visited on 09/03/2025).
- [15] Ione Fine and Geoffrey M. Boynton. “Pulse trains to percepts: the challenge of creating a perceptually intelligible world with sight recovery technologies”. en. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1677 (Sept. 2015), p. 20140208. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2014.

0208. URL: <https://royalsocietypublishing.org/doi/10.1098/rstb.2014.0208> (visited on 02/03/2023).
- [16] Michael Beyeler et al. “Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies”. en. In: *J. Neural Eng.* (2017), p. 20.
- [17] Michael Beyeler et al. “A model of ganglion axon pathways accounts for percepts elicited by retinal implants”. en. In: *Scientific Reports* 9.1 (June 2019), pp. 1–16. ISSN: 2045-2322. DOI: 10.1038/s41598-019-45416-4. URL: <https://www.nature.com/articles/s41598-019-45416-4> (visited on 10/16/2019).
- [18] Eduardo Fernández et al. “Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex”. en. In: *Journal of Clinical Investigation* 131.23 (Dec. 2021), e151331. ISSN: 1558-8238. DOI: 10.1172/JCI151331. URL: <https://www.jci.org/articles/view/151331> (visited on 09/26/2024).
- [19] Michael P Barry et al. “Preliminary visual function for the first human with the Intracortical Visual Prosthesis (ICVP)”. In: *Investigative Ophthalmology & Visual Science* 64.8 (June 2023), p. 2842. ISSN: 1552-5783.
- [20] Lucas Nadolskis et al. “Aligning Visual Prosthetic Development With Implantee Needs”. In: *Translational Vision Science & Technology* 13.11 (Nov. 2024), p. 28. ISSN: 2164-2591. DOI: 10.1167/tvst.13.11.28. URL: <https://doi.org/10.1167/tvst.13.11.28> (visited on 09/01/2025).
- [21] Michael Beyeler and Melani Sanchez Garcia. *Towards a Smart Bionic Eye: AI-Powered Artificial Vision for the Treatment of Incurable Blindness*. en-us. June 2022. DOI: 10.31219/osf.io/fpdyv. URL: <https://osf.io/fpdyv/> (visited on 06/21/2022).

- [22] Jacob Granley and Michael Beyeler. “A Computational Model of Phosphene Appearance for Epiretinal Prostheses”. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. ISSN: 2694-0604. Nov. 2021, pp. 4477–4481. DOI: 10.1109/EMBC46164.2021.9629663. URL: <https://ieeexplore.ieee.org/document/9629663/?arnumber=9629663> (visited on 01/09/2025).
- [23] Eirini Schoinas et al. “Evaluating Deep Human-in-the-Loop Optimization for Retinal Implants Using Sighted Participants”. In: *2025 47th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. arXiv:2502.00177 [cs]. IEEE, Apr. 2025. DOI: 10.48550/arXiv.2502.00177. URL: <http://arxiv.org/abs/2502.00177> (visited on 09/01/2025).
- [24] Jacob Granley, Lucas Relic, and Michael Beyeler. “Hybrid Neural Autoencoders for Stimulus Encoding in Visual and Other Sensory Neuroprostheses”. en. In: *Advances in Neural Information Processing Systems* 35 (Dec. 2022), pp. 22671–22685. URL: https://proceedings.neurips.cc/paper_files/paper/2022/hash/8e9a6582caa59fda0302349702965171-Abstract-Conference.html (visited on 12/02/2024).
- [25] Jacob Granley et al. “Human-in-the-Loop Optimization for Deep Stimulus Encoding in Visual Prostheses”. en. In: *Advances in Neural Information Processing Systems* 36 (Dec. 2023), pp. 79376–79398. URL: https://proceedings.neurips.cc/paper_files/paper/2023/hash/fb06bc3abcece7b8725a8b83b8fa3632-Abstract-Conference.html (visited on 01/09/2025).
- [26] Dariusz Man and Ryszard Olchawa. “Brain Biophysics: Perception, Consciousness, Creativity. Brain Computer Interface (BCI)”. en. In: *Biomedical Engineering and Neuroscience*. Ed. by Wojciech P. Hunek and Szczepan Paszkiel. Cham:

- Springer International Publishing, 2018, pp. 38–44. ISBN: 978-3-319-75025-5. DOI: 10.1007/978-3-319-75025-5_5.
- [27] Jon H. Kaas. “Why Does the Brain Have So Many Visual Areas?” In: *Journal of Cognitive Neuroscience* 1.2 (Apr. 1989), pp. 121–135. ISSN: 0898-929X. DOI: 10.1162/jocn.1989.1.2.121. URL: <https://doi.org/10.1162/jocn.1989.1.2.121> (visited on 09/02/2025).
- [28] Eberhart Zrenner. “Fighting Blindness with Microelectronics”. In: *Science Translational Medicine* 5.210 (Nov. 2013). Publisher: American Association for the Advancement of Science, 210ps16–210ps16. DOI: 10.1126/scitranslmed.3007399. URL: <https://www.science.org/doi/10.1126/scitranslmed.3007399> (visited on 09/05/2025).
- [29] Jonathan J. Nassi and Edward M. Callaway. “Parallel processing strategies of the primate visual system”. en. In: *Nature Reviews Neuroscience* 10.5 (May 2009). Publisher: Nature Publishing Group, pp. 360–372. ISSN: 1471-0048. DOI: 10.1038/nrn2619. URL: <https://www.nature.com/articles/nrn2619> (visited on 09/06/2025).
- [30] David C. Van Essen and John H. R. Maunsell. “Hierarchical organization and functional streams in the visual cortex”. In: *Trends in Neurosciences* 6 (Jan. 1983), pp. 370–375. ISSN: 0166-2236. DOI: 10.1016/0166-2236(83)90167-4. URL: <https://www.sciencedirect.com/science/article/pii/0166223683901674> (visited on 09/02/2025).
- [31] Rupert R A Bourne et al. “Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe in 2015: magnitude, temporal trends and projections”. In: *The British Journal of Ophthalmology* 102.5 (May 2018), pp. 575–585. ISSN: 0007-1161. DOI: 10.1136/bjophthalmol-2017-311258. URL:

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5909755/> (visited on 09/02/2025).
- [32] Param Vaidya. “Retinitis Pigmentosa: Disease Encumbrance in the Eurozone”. In: *International Journal of Ophthalmology and Clinical Research* (2015). URL: https://clinmedjournals.org/articles/ijocr/ijocr-2-030.php?jid=ijocr&utm_source=chatgpt.com (visited on 09/02/2025).
- [33] Rupert R A Bourne et al. “Global Prevalence of Blindness and Distance and Near Vision Impairment in 2020: progress towards the Vision 2020 targets and what the future holds.” In: *Investigative Ophthalmology & Visual Science* 61.7 (June 2020), p. 2317. ISSN: 1552-5783.
- [34] Rupert Bourne. “Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis - The Lancet Global Health”. In: *The Lancet* 5.9 (2017). URL: [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(17\)30293-0/fulltext?uuid=ezRypWhIesimRf2q0082](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(17)30293-0/fulltext?uuid=ezRypWhIesimRf2q0082) (visited on 09/03/2025).
- [35] Seth R Flaxman et al. “Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis”. In: *The Lancet Global Health* 5.12 (Dec. 2017), e1221–e1234. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(17)30393-5. URL: <https://www.sciencedirect.com/science/article/pii/S2214109X17303935> (visited on 09/03/2025).
- [36] Yuchen Liu et al. “Gene Therapy for Retinitis Pigmentosa: Current Challenges and New Progress”. en. In: *Biomolecules* 14.8 (Aug. 2024). Publisher: Multidisciplinary Digital Publishing Institute, p. 903. ISSN: 2218-273X. DOI: 10.3390/biom14080903. URL: <https://www.mdpi.com/2218-273X/14/8/903> (visited on 09/03/2025).

- [37] Peter A. Campochiaro et al. “Oral *N*-acetylcysteine improves cone function in retinitis pigmentosa patients in phase I trial”. en. In: *The Journal of Clinical Investigation* 130.3 (Mar. 2020). Publisher: American Society for Clinical Investigation, pp. 1527–1541. ISSN: 0021-9738. DOI: 10.1172/JCI132990. URL: https://www.jci.org/articles/view/132990?utm_source=chatgpt.com (visited on 09/03/2025).
- [38] Andrzej T. Foik et al. “Detailed Visual Cortical Responses Generated by Retinal Sheet Transplants in Rats with Severe Retinal Degeneration”. en. In: *Journal of Neuroscience* 38.50 (Dec. 2018). Publisher: Society for Neuroscience Section: Research Articles, pp. 10709–10724. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1279-18.2018. URL: <https://www.jneurosci.org/content/38/50/10709> (visited on 09/05/2025).
- [39] José-Alain Sahel et al. “Partial recovery of visual function in a blind patient after optogenetic therapy”. en. In: *Nature Medicine* 27.7 (July 2021). Publisher: Nature Publishing Group, pp. 1223–1229. ISSN: 1546-170X. DOI: 10.1038/s41591-021-01351-4. URL: <https://www.nature.com/articles/s41591-021-01351-4> (visited on 09/03/2025).
- [40] Jeffrey H. Stern et al. “Regenerating Eye Tissues to Preserve and Restore Vision”. English. In: *Cell Stem Cell* 22.6 (June 2018). Publisher: Elsevier, pp. 834–849. ISSN: 1934-5909, 1875-9777. DOI: 10.1016/j.stem.2018.05.013. URL: [https://www.cell.com/cell-stem-cell/abstract/S1934-5909\(18\)30231-5](https://www.cell.com/cell-stem-cell/abstract/S1934-5909(18)30231-5) (visited on 09/03/2025).
- [41] Botond Roska and José-Alain Sahel. “Restoring vision”. en. In: *Nature* 557.7705 (May 2018). Publisher: Nature Publishing Group, pp. 359–367. ISSN: 1476-4687.

- DOI: 10.1038/s41586-018-0076-4. URL: <https://www.nature.com/articles/s41586-018-0076-4> (visited on 09/05/2025).
- [42] Robert K. Shepherd et al. “Visual prostheses for the blind”. en. In: *Trends in Biotechnology* 31.10 (Oct. 2013), pp. 562–571. ISSN: 0167-7799. DOI: 10.1016/j.tibtech.2013.07.001. URL: <https://www.sciencedirect.com/science/article/pii/S0167779913001595> (visited on 04/05/2021).
- [43] G. S. Brindley and W. S. Lewin. “The sensations produced by electrical stimulation of the visual cortex”. en. In: *The Journal of Physiology* 196.2 (1968), pp. 479–493. ISSN: 1469-7793. DOI: 10.1113/jphysiol.1968.sp008519. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1968.sp008519> (visited on 09/01/2025).
- [44] Wm H. Dobelle. “Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex”. en-US. In: *ASAIO Journal* 46.1 (Feb. 2000), p. 3. ISSN: 1058-2916. URL: https://journals.lww.com/asaiojournal/fulltext/2000/01000/Artificial_Vision_for_the_Blind_by_Connecting_a.2.aspx (visited on 03/16/2025).
- [45] Yvonne Hsu-Lin Luo and Lyndon da Cruz. “The Argus® II Retinal Prosthesis System”. In: *Progress in Retinal and Eye Research* 50 (Jan. 2016), pp. 89–107. ISSN: 1350-9462. DOI: 10.1016/j.preteyeres.2015.09.003. URL: <https://www.sciencedirect.com/science/article/pii/S1350946215000701> (visited on 12/02/2024).
- [46] Lauren N. Ayton et al. “First-in-Human Trial of a Novel Suprachoroidal Retinal Prosthesis”. en. In: *PLoS ONE* 9.12 (Dec. 2014). Ed. by Keisuke Mori, e115239. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0115239. URL: <https://dx.plos.org/10.1371/journal.pone.0115239> (visited on 12/02/2024).

- [47] Katarina Stingl et al. “Subretinal Visual Implant Alpha IMS – Clinical trial interim report”. In: *Vision Research*. Sight restoration: prosthetics, optogenetics and gene therapy 111 (June 2015), pp. 149–160. ISSN: 0042-6989. DOI: 10.1016/j.visres.2015.03.001. URL: <https://www.sciencedirect.com/science/article/pii/S0042698915000784> (visited on 12/02/2024).
- [48] D. Palanker et al. “Simultaneous perception of prosthetic and natural vision in AMD patients”. en. In: *Nature Communications* 13.1 (Jan. 2022). Publisher: Nature Publishing Group, p. 513. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28125-x. URL: <https://www.nature.com/articles/s41467-022-28125-x> (visited on 12/02/2024).
- [49] Claude Veraart et al. “Pattern Recognition with the Optic Nerve Visual Prosthesis”. en. In: *Artificial Organs* 27.11 (2003), pp. 996–1004. ISSN: 1525-1594. DOI: 10.1046/j.1525-1594.2003.07305.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1525-1594.2003.07305.x> (visited on 12/02/2024).
- [50] Fivos Panetsos et al. “Consistent Phosphenes Generated by Electrical Microstimulation of the Visual Thalamus. An Experimental Approach for Thalamic Visual Neuroprostheses”. English. In: *Frontiers in Neuroscience* 5 (July 2011). Publisher: Frontiers. ISSN: 1662-453X. DOI: 10.3389/fnins.2011.00084. URL: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2011.00084/full> (visited on 12/02/2024).
- [51] Elon Musk and Neuralink. “An Integrated Brain-Machine Interface Platform With Thousands of Channels”. EN. In: *Journal of Medical Internet Research* 21.10 (Oct. 2019). Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto,

- Canada, e16194. DOI: 10.2196/16194. URL: <https://www.jmir.org/2019/10/e16194> (visited on 12/02/2024).
- [52] Michael S. Beauchamp et al. “Dynamic Stimulation of Visual Cortex Produces Form Vision in Sighted and Blind Humans”. In: *Cell* 181.4 (May 2020), 774–783.e5. ISSN: 0092-8674. DOI: 10.1016/j.cell.2020.04.033. URL: <https://www.sciencedirect.com/science/article/pii/S0092867420304967> (visited on 09/26/2024).
- [53] *Bionic Vision - Advancing Sight Restoration*. URL: <https://www.bionic-vision.org/implants> (visited on 09/03/2025).
- [54] Edwin M. Maynard, Craig T. Nordhausen, and Richard A. Normann. “The Utah Intracortical Electrode Array: A recording structure for potential brain-computer interfaces”. In: *Electroencephalography and Clinical Neurophysiology* 102.3 (Mar. 1997), pp. 228–239. ISSN: 0013-4694. DOI: 10.1016/S0013-4694(96)95176-0. URL: <https://www.sciencedirect.com/science/article/pii/S0013469496951760> (visited on 09/03/2025).
- [55] Larry Greenemeier. “FDA Approves First Retinal Implant”. en. In: *Nature* (Feb. 2013). Publisher: Nature Publishing Group. ISSN: 1476-4687. DOI: 10.1038/nature.2013.12439. URL: <https://www.nature.com/articles/nature.2013.12439> (visited on 09/03/2025).
- [56] A. K. Ahuja et al. “Factors Affecting Perceptual Threshold in Argus II Retinal Prosthesis Subjects”. In: *Translational Vision Science & Technology* 2.4 (Apr. 2013), p. 1. ISSN: 2164-2591. DOI: 10.1167/tvst.2.4.1. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3763895/> (visited on 09/03/2025).
- [57] Lan Yue et al. “Retina-electrode interface properties and vision restoration by two generations of retinal prostheses in one patient-one in each eye”. eng. In:

- Journal of Neural Engineering* 17.2 (Apr. 2020), p. 026020. ISSN: 1741-2552. DOI: 10.1088/1741-2552/ab7c8f.
- [58] Eduardo Fernandez. *Pilot Study for the Development of a Cortical Visual Neuroprosthesis for the Blind Based on Intracortical Microelectrodes*. Clinical trial registration NCT02983370. submitted: 2016-09-12. clinicaltrials.gov, May 2024. URL: <https://clinicaltrials.gov/study/NCT02983370> (visited on 09/03/2025).
- [59] Cordelia Erickson-Davis and Helma Korzybska. *What do blind people “see” with retinal prostheses? Observations and qualitative reports of epiretinal implant users*. en. preprint. Neuroscience, Feb. 2020. DOI: 10.1101/2020.02.03.932905. URL: <http://biorxiv.org/lookup/doi/10.1101/2020.02.03.932905> (visited on 02/03/2023).
- [60] Nadia Paraskevoudi and John S. Pezaris. “Eye Movement Compensation and Spatial Updating in Visual Prosthetics: Mechanisms, Limitations and Future Directions”. English. In: *Frontiers in Systems Neuroscience* 12 (Feb. 2019). Publisher: Frontiers. ISSN: 1662-5137. DOI: 10.3389/fnsys.2018.00073. URL: <https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2018.00073/full> (visited on 09/03/2025).
- [61] Angélica Pérez Fornos et al. “Temporal Properties of Visual Perception on Electrical Stimulation of the Retina”. In: *Investigative Ophthalmology & Visual Science* 53.6 (May 2012), pp. 2720–2731. ISSN: 1552-5783. DOI: 10.1167/iovs.11-9344. URL: <https://doi.org/10.1167/iovs.11-9344> (visited on 11/08/2022).
- [62] Spencer C. Chen et al. “Simulating prosthetic vision: I. Visual models of phosphenes”. In: *Vision Research* 49.12 (June 2009), pp. 1493–1506. ISSN: 0042-6989. DOI: 10.1016/j.visres.2009.02.003. URL: <https://www.sciencedirect.com/science/article/pii/S0042698909000467> (visited on 09/03/2025).

- [63] Yuchen Hou et al. *Axonal stimulation affects the linear summation of single-point perception in three Argus II users*. en. preprint. Ophthalmology, July 2023. DOI: 10.1101/2023.07.21.23292908. URL: <http://medrxiv.org/lookup/doi/10.1101/2023.07.21.23292908> (visited on 10/24/2023).
- [64] William H. Bosking et al. “Percepts evoked by multi-electrode stimulation of human visual cortex”. en. In: *Brain Stimulation* (Aug. 2022). ISSN: 1935-861X. DOI: 10.1016/j.brs.2022.08.007. URL: <https://www.sciencedirect.com/science/article/pii/S1935861X22001851> (visited on 08/19/2022).
- [65] Alan Horsager, Robert J. Greenberg, and Ione Fine. “Spatiotemporal Interactions in Retinal Prosthesis Subjects”. en. In: *Investigative Ophthalmology & Visual Science* 51.2 (Feb. 2010), p. 1223. ISSN: 1552-5783. DOI: 10.1167/iovs.09-3746. URL: <http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.09-3746> (visited on 02/03/2023).
- [66] David Youssef et al. “Neuronal Spiking Responses to Direct Electrical Microstimulation in the Human Cortex”. en. In: *Journal of Neuroscience* 43.24 (June 2023). Publisher: Society for Neuroscience Section: Research Articles, pp. 4448–4460. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1666-22.2023. URL: <https://www.jneurosci.org/content/43/24/4448> (visited on 02/10/2025).
- [67] Karthik Kumaravelu and Warren M. Grill. “Neural mechanisms of the temporal response of cortical neurons to intracortical microstimulation”. In: *Brain Stimulation* 17.2 (Mar. 2024), pp. 365–381. ISSN: 1935-861X. DOI: 10.1016/j.brs.2024.03.012. URL: <https://www.sciencedirect.com/science/article/pii/S1935861X24000482> (visited on 02/10/2025).
- [68] Ione Fine and Geoffrey M. Boynton. “A virtual patient simulation modeling the neural and perceptual effects of human visual cortical stimulation, from pulse

- trains to percepts”. en. In: *Scientific Reports* 14.1 (July 2024). Publisher: Nature Publishing Group, p. 17400. ISSN: 2045-2322. DOI: 10.1038/s41598-024-65337-1. URL: <https://www.nature.com/articles/s41598-024-65337-1> (visited on 09/03/2025).
- [69] Justin Kasowski and Michael Beyeler. “Immersive Virtual Reality Simulations of Bionic Vision”. In: *Augmented Humans 2022*. AHs 2022. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 82–93. ISBN: 978-1-4503-9632-5. DOI: 10.1145/3519391.3522752. URL: <https://doi.org/10.1145/3519391.3522752> (visited on 10/06/2022).
- [70] Michael Beyeler. “Biophysical model of axonal stimulation in epiretinal visual prostheses”. In: *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. ISSN: 1948-3554. Mar. 2019, pp. 348–351. DOI: 10.1109/NER.2019.8716969. URL: <https://ieeexplore.ieee.org/document/8716969/> (visited on 08/26/2025).
- [71] Aman S Aberra, Angel V Peterchev, and Warren M Grill. “Biophysically realistic neuron models for simulation of cortical stimulation”. en. In: *Journal of Neural Engineering* 15.6 (Oct. 2018). Publisher: IOP Publishing, p. 066023. ISSN: 1741-2552. DOI: 10.1088/1741-2552/aadbb1. URL: <https://dx.doi.org/10.1088/1741-2552/aadbb1> (visited on 09/03/2025).
- [72] Aiwen Xu and Michael Beyeler. “Retinal ganglion cells undergo cell type—specific functional changes in a computational model of cone-mediated retinal degeneration”. English. In: *Frontiers in Neuroscience* 17 (May 2023). Publisher: Frontiers. ISSN: 1662-453X. DOI: 10.3389/fnins.2023.1147729. URL: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1147729/full> (visited on 09/03/2025).

- [73] Fabrizio Grani. “Improving cortical visual prostheses using closed-loop stimulation approaches”. en. Pages: 1. <http://purl.org/dc/dcmitype/Text>. Universidad Miguel Hernández de Elche, 2024. URL: <https://dialnet.unirioja.es/servlet/tesis?codigo=376066> (visited on 08/26/2025).
- [74] M. Beyeler et al. “pulse2percept: A Python-based simulation framework for bionic vision”. In: *Proceedings of the 16th Science in Python Conference*. Ed. by K. Huff et al. 2017, pp. 81–88. DOI: 10.25080/shinma-7f4c6e7-00c.
- [75] Robert W. Thompson Jr et al. “Facial Recognition Using Simulated Prosthetic Pixelized Vision”. In: *Investigative Ophthalmology & Visual Science* 44.11 (Nov. 2003), pp. 5035–5042. ISSN: 1552-5783. DOI: 10.1167/iovs.03-0341. URL: <https://doi.org/10.1167/iovs.03-0341> (visited on 09/04/2025).
- [76] Alan Horsager et al. “Predicting Visual Sensitivity in Retinal Prosthesis Patients”. en. In: *Investigative Ophthalmology & Visual Science* 50.4 (Apr. 2009), p. 1483. ISSN: 1552-5783. DOI: 10.1167/iovs.08-2595. URL: <http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.08-2595> (visited on 02/03/2023).
- [77] Devyani Nanduri et al. “Frequency and Amplitude Modulation Have Different Effects on the Percepts Elicited by Retinal Stimulation”. In: *Investigative Ophthalmology & Visual Science* 53.1 (Jan. 2012), pp. 205–214. ISSN: 1552-5783. DOI: 10.1167/iovs.11-8401. URL: <https://doi.org/10.1167/iovs.11-8401> (visited on 11/22/2021).
- [78] David Avraham et al. “Retinal prosthetic vision simulation: temporal aspects”. en. In: *Journal of Neural Engineering* (2021). ISSN: 1741-2552. DOI: 10.1088/1741-2552/ac1b6c. URL: <http://iopscience.iop.org/article/10.1088/1741-2552/ac1b6c> (visited on 08/07/2021).

- [79] Justin Michael Kasowski. “Virtual Reality Simulations for the Advancement of Visual Prosthetics”. English. ISBN: 9798381687101. Ph.D. United States – California: University of California, Santa Barbara, 2023. URL: <https://www.proquest.com/docview/2923558720/abstract/E401E93DE5EC4A3BPQ/1> (visited on 09/03/2025).
- [80] Yuchen Hou et al. “Predicting the Temporal Dynamics of Prosthetic Vision”. In: *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. ISSN: 2694-0604. July 2024, pp. 1–4. DOI: 10.1109/EMBC53108.2024.10782668. URL: <https://ieeexplore.ieee.org/abstract/document/10782668> (visited on 09/03/2025).
- [81] Simone Romeni et al. “A computational model to design wide field-of-view optic nerve neuroprostheses”. In: *iScience* 27.12 (Dec. 2024), p. 111321. ISSN: 2589-0042. DOI: 10.1016/j.isci.2024.111321. URL: <https://www.sciencedirect.com/science/article/pii/S258900422402546X> (visited on 09/04/2025).
- [82] Jaap de Ruyter van Steveninck et al. “End-to-end optimization of prosthetic vision”. In: *Journal of Vision* 22.2 (Feb. 2022), p. 20. ISSN: 1534-7362. DOI: 10.1167/jov.22.2.20. URL: <https://doi.org/10.1167/jov.22.2.20> (visited on 12/02/2024).
- [83] Jacob Granley, Alexander Riedel, and Michael Beyeler. “Adapting Brain-Like Neural Networks for Modeling Cortical Visual Prostheses”. en. In: Oct. 2022. URL: <https://openreview.net/forum?id=x0zleacdnee> (visited on 09/04/2025).
- [84] Haozhe Zac Wang and Yan Tat Wong. “A novel simulation paradigm utilising MRI-derived phosphene maps for cortical prosthetic vision”. en. In: *Journal of Neural Engineering* 20.4 (Aug. 2023). Publisher: IOP Publishing, p. 046027. ISSN:

BIBLIOGRAPHY

- 1741-2552. DOI: 10.1088/1741-2552/aceca2. URL: <https://dx.doi.org/10.1088/1741-2552/aceca2> (visited on 09/03/2025).
- [85] Maureen Van Der Grinten et al. *Biologically plausible phosphene simulation for the differentiable optimization of visual cortical prostheses*. en. preprint. Neuroscience, Dec. 2022. DOI: 10.1101/2022.12.23.521749. URL: <http://biorxiv.org/lookup/doi/10.1101/2022.12.23.521749> (visited on 05/17/2023).
- [86] Christine A. Curcio et al. “Human photoreceptor topography”. en. In: *Journal of Comparative Neurology* 292.4 (1990), pp. 497–523. ISSN: 1096-9861. DOI: 10.1002/cne.902920402. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.902920402> (visited on 09/03/2025).
- [87] Andrew B. Watson. “A formula for human retinal ganglion cell receptive field density as a function of visual field location”. In: *Journal of Vision* 14.7 (June 2014), p. 15. ISSN: 1534-7362. DOI: 10.1167/14.7.15. URL: <https://doi.org/10.1167/14.7.15> (visited on 09/03/2025).
- [88] J. R. Polimeni, M. Balasubramanian, and E. L. Schwartz. “Multi-area visuotopic map complexes in macaque striate and extra-striate cortex”. In: *Vision Research* 46.20 (Oct. 2006), pp. 3336–3359. ISSN: 0042-6989. DOI: 10.1016/j.visres.2006.03.006. URL: <https://www.sciencedirect.com/science/article/pii/S0042698906001428> (visited on 09/03/2025).
- [89] Noah C. Benson et al. “The Retinotopic Organization of Striate Cortex Is Well Predicted by Surface Topology”. In: *Current Biology* 22.21 (Nov. 2012), pp. 2081–2085. ISSN: 0960-9822. DOI: 10.1016/j.cub.2012.09.014. URL: <https://www.sciencedirect.com/science/article/pii/S0960982212010779> (visited on 09/03/2025).

- [90] Nicole Han et al. “Deep Learning–Based Scene Simplification for Bionic Vision”. In: *arXiv:2102.00297 [cs]* (Jan. 2021). arXiv: 2102.00297. URL: <http://arxiv.org/abs/2102.00297> (visited on 03/22/2021).
- [91] Wei Tong et al. “Stimulation Strategies for Improving the Resolution of Retinal Prostheses”. English. In: *Frontiers in Neuroscience* 14 (Mar. 2020). Publisher: Frontiers. ISSN: 1662-453X. DOI: 10.3389/fnins.2020.00262. URL: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2020.00262/full> (visited on 03/16/2025).
- [92] Gerald Dumm et al. “Virtual Electrodes by Current Steering in Retinal Prostheses”. In: *Investigative Ophthalmology & Visual Science* 55.12 (Dec. 2014), pp. 8077–8085. ISSN: 1552-5783. DOI: 10.1167/iovs.14-15391. URL: <https://doi.org/10.1167/iovs.14-15391> (visited on 09/04/2025).
- [93] Dorsa Haji Ghaffari et al. “Closed-loop Optimization of Retinal Ganglion Cell Responses to Epiretinal Stimulation: A Computational Study”. In: *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. ISSN: 1948-3554. May 2021, pp. 597–600. DOI: 10.1109/NER49283.2021.9441437. URL: <https://ieeexplore.ieee.org/document/9441437/?arnumber=9441437> (visited on 10/15/2024).
- [94] Nishal P. Shah et al. “Optimization of Electrical Stimulation for a High-Fidelity Artificial Retina”. In: *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. ISSN: 1948-3554. Mar. 2019, pp. 714–718. DOI: 10.1109/NER.2019.8716987. URL: <https://ieeexplore.ieee.org/abstract/document/8716987> (visited on 01/09/2025).
- [95] Nishal Pradeepbhai Shah et al. “Precise control of neural activity using dynamically optimized electrical stimulation”. In: *eLife* 13 (Nov. 2024). Ed. by Michael Beyeler,

- Joshua I Gold, and Mohit Shivdasani. Publisher: eLife Sciences Publications, Ltd, e83424. ISSN: 2050-084X. DOI: 10.7554/eLife.83424. URL: <https://doi.org/10.7554/eLife.83424> (visited on 09/04/2025).
- [96] Martin J Spencer et al. “Global activity shaping strategies for a retinal implant”. en. In: *Journal of Neural Engineering* 16.2 (Jan. 2019). Publisher: IOP Publishing, p. 026008. ISSN: 1741-2552. DOI: 10.1088/1741-2552/aaf071. URL: <https://dx.doi.org/10.1088/1741-2552/aaf071> (visited on 01/09/2025).
- [97] Tristan Fauvel and Matthew Chalk. *Human-in-the-loop optimization of visual prosthetic stimulation*. en. Tech. rep. 4. Nov. 2021, p. 2021.11.24.469867. URL: <https://www.biorxiv.org/content/10.1101/2021.11.24.469867v1> (visited on 11/26/2021).
- [98] Simone Romeni, Davide Zoccolan, and Silvestro Micera. “A machine learning framework to optimize optic nerve electrical stimulation for vision restoration”. In: *Patterns* 2.7 (June 2021), p. 100286. ISSN: 2666-3899. DOI: 10.1016/j.patter.2021.100286. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8276026/> (visited on 05/17/2023).
- [99] Maureen van der Grinten et al. “Towards biologically plausible phosphene simulation for the differentiable optimization of visual cortical prostheses”. In: *eLife* 13 (Feb. 2024). Ed. by Chris I Baker and Michael P Barry. Publisher: eLife Sciences Publications, Ltd, e85812. ISSN: 2050-084X. DOI: 10.7554/eLife.85812. URL: <https://doi.org/10.7554/eLife.85812> (visited on 09/04/2025).
- [100] Lucas Relic et al. “Deep Learning–Based Perceptual Stimulus Encoder for Bionic Vision”. In: *Proceedings of the Augmented Humans International Conference 2022*. AHs ’22. New York, NY, USA: Association for Computing Machinery, Apr. 2022,

- pp. 323–325. ISBN: 978-1-4503-9632-5. DOI: 10.1145/3519391.3524034. URL: <https://dl.acm.org/doi/10.1145/3519391.3524034> (visited on 01/09/2025).
- [101] Burcu Küçükoğlu et al. “Optimization of Neuroprosthetic Vision via End-to-End Deep Reinforcement Learning”. en. In: *International Journal of Neural Systems* 32.11 (Nov. 2022), p. 2250052. ISSN: 0129-0657, 1793-6462. DOI: 10.1142/S0129065722500526. URL: <https://www.worldscientific.com/doi/10.1142/S0129065722500526> (visited on 09/04/2025).
- [102] Burcu Küçükoğlu et al. *End-to-end Learning of Safe Stimulation Parameters for Cortical Neuroprosthetic Vision*. en. Pages: 2025.01.23.634543 Section: New Results. Jan. 2025. DOI: 10.1101/2025.01.23.634543. URL: <https://www.biorxiv.org/content/10.1101/2025.01.23.634543v1> (visited on 02/04/2025).
- [103] Yuli Wu et al. “A Deep Learning-based in silico Framework for Optimization on Retinal Prosthetic Stimulation”. In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. ISSN: 2694-0604. July 2023, pp. 1–4. DOI: 10.1109/EMBC40787.2023.10340288. URL: <https://ieeexplore.ieee.org/abstract/document/10340288> (visited on 09/04/2025).
- [104] Yuli Wu et al. *Optimizing Retinal Prosthetic Stimuli with Conditional Invertible Neural Networks*. arXiv:2403.04884 [cs]. July 2024. DOI: 10.48550/arXiv.2403.04884. URL: <http://arxiv.org/abs/2403.04884> (visited on 09/04/2025).
- [105] Yuli Wu et al. *Visual Fixation-Based Retinal Prosthetic Simulation*. arXiv:2410.11688 [cs]. Oct. 2024. DOI: 10.48550/arXiv.2410.11688. URL: <http://arxiv.org/abs/2410.11688> (visited on 09/04/2025).
- [106] Chuanqing Wang et al. “Exploring Effective Stimulus Encoding via Vision System Modeling for Visual Prostheses”. en. In: Oct. 2023. URL: <https://openreview.net/forum?id=cKAUvMePUN> (visited on 09/04/2025).

- [107] Domingos Castro et al. “Neural activity shaping in visual prostheses with deep learning”. en. In: *Journal of Neural Engineering* 21.4 (July 2024). Publisher: IOP Publishing, p. 046025. ISSN: 1741-2552. DOI: 10.1088/1741-2552/ad6186. URL: <https://dx.doi.org/10.1088/1741-2552/ad6186> (visited on 09/04/2025).
- [108] Jacob Lavoie et al. “Learning to see via epiretinal implant stimulation in silico with model-based deep reinforcement learning”. en. In: *Biomedical Physics & Engineering Express* 10.2 (Jan. 2024). Publisher: IOP Publishing, p. 025006. ISSN: 2057-1976. DOI: 10.1088/2057-1976/acf1a5. URL: <https://dx.doi.org/10.1088/2057-1976/acf1a5> (visited on 09/04/2025).
- [109] Franklin Leong et al. “An actor-model framework for visual sensory encoding”. en. In: *Nature Communications* 15.1 (Jan. 2024). Publisher: Nature Publishing Group, p. 808. ISSN: 2041-1723. DOI: 10.1038/s41467-024-45105-5. URL: <https://www.nature.com/articles/s41467-024-45105-5> (visited on 04/02/2024).
- [110] Lauren H. Jepson et al. “High-Fidelity Reproduction of Spatiotemporal Visual Signals for Retinal Prosthesis”. English. In: *Neuron* 83.1 (July 2014). Publisher: Elsevier, pp. 87–92. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2014.04.044. URL: [https://www.cell.com/neuron/abstract/S0896-6273\(14\)00359-6](https://www.cell.com/neuron/abstract/S0896-6273(14)00359-6) (visited on 09/05/2025).
- [111] Martin J. Spencer et al. “Neural activity shaping utilizing a partitioned target pattern”. en. In: *Journal of Neural Engineering* 18 (Mar. 2021). 5, p. 046025. ISSN: 1741-2552. DOI: 10.1088/1741-2552/abecc4. URL: <https://doi.org/10.1088/1741-2552/abecc4> (visited on 10/06/2021).
- [112] Domingos Castro et al. *Neural Activity Shaping in Visual Prostheses with Deep Learning*. en. Pages: 2023.12.20.572123 Section: New Results. Dec. 2023. DOI:

- 10.1101/2023.12.20.572123. URL: <https://www.biorxiv.org/content/10.1101/2023.12.20.572123v1> (visited on 09/26/2024).
- [113] Yuki Minai et al. “MiSO: Optimizing brain stimulation to create neural population activity states”. en. In: *Advances in Neural Information Processing Systems* 37 (Dec. 2024), pp. 24126–24149. URL: https://proceedings.neurips.cc/paper_files/paper/2024/hash/2af641762dc02035c31a9314b2d090b6-Abstract-Conference.html (visited on 08/26/2025).
- [114] Fotios Drakopoulos and Sarah Verhulst. “A Differentiable Optimisation Framework for The Design of Individualised DNN-based Hearing-Aid Strategies”. In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. ISSN: 2379-190X. May 2022, pp. 351–355. DOI: 10.1109/ICASSP43922.2022.9747683.
- [115] Fotios Drakopoulos, Arthur Van Den Broucke, and Sarah Verhulst. “A DNN-Based Hearing-Aid Strategy For Real-Time Processing: One Size Fits All”. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. June 2023, pp. 1–5. DOI: 10.1109/ICASSP49357.2023.10094887.
- [116] K. Stingl et al. “Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS”. In: *Proc Biol Sci* 280.1757 (Apr. 2013), p. 20130077. ISSN: 1471-2954 (Electronic) 0962-8452 (Linking). DOI: 10.1098/rspb.2013.0077.
- [117] Henri Lorach et al. “Photovoltaic restoration of sight with high visual acuity”. en. In: *Nature Medicine* 21.5 (May 2015), pp. 476–482. ISSN: 1078-8956, 1546-170X. DOI: 10.1038/nm.3851. URL: <https://www.nature.com/articles/nm.3851> (visited on 08/26/2025).

- [118] Y. H. Luo and L. da Cruz. “The Argus((R)) II Retinal Prosthesis System”. In: *Prog Retin Eye Res* 50 (Jan. 2016), pp. 89–107. ISSN: 1873-1635 (Electronic) 1350-9462 (Linking). DOI: 10.1016/j.preteyeres.2015.09.003.
- [119] Lauren N. Ayton et al. “An update on retinal prostheses”. en. In: *Clinical Neurophysiology* 131.6 (June 2020). 1, pp. 1383–1398. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2019.11.029. URL: <https://www.sciencedirect.com/science/article/pii/S1388245719313264> (visited on 10/04/2022).
- [120] Yvonne H-L. Luo et al. “Long-term Repeatability and Reproducibility of Phosphene Characteristics in Chronically Implanted Argus II Retinal Prosthesis Subjects”. en. In: *American Journal of Ophthalmology* 170 (Oct. 2016), pp. 100–109. ISSN: 0002-9394. DOI: 10.1016/j.ajo.2016.07.021. URL: <https://www.sciencedirect.com/science/article/pii/S0002939416303671> (visited on 08/08/2023).
- [121] Bryan W. Jones and Robert E. Marc. “Retinal remodeling during retinal degeneration”. en. In: *Experimental Eye Research* 81.2 (Aug. 2005), pp. 123–137. ISSN: 0014-4835. DOI: 10.1016/j.exer.2005.03.006. URL: <https://www.sciencedirect.com/science/article/pii/S0014483505000989> (visited on 05/03/2021).
- [122] Kathleen E. Finn et al. “A Patient-Specific Computational Framework for the Argus II Implant”. In: *IEEE Open Journal of Engineering in Medicine and Biology* 1 (2020), pp. 190–196. ISSN: 2644-1276. DOI: 10.1109/OJEMB.2020.3001563. URL: <https://ieeexplore.ieee.org/document/9115241/> (visited on 08/26/2025).
- [123] Andrew C. Weitz et al. “Improving the spatial resolution of epiretinal implants by increasing stimulus pulse duration”. en. In: *Science Translational Medicine* 7.318 (Dec. 2015). ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.aac4877.

- URL: <https://www.science.org/doi/10.1126/scitranslmed.aac4877> (visited on 05/17/2023).
- [124] Laura Ferlauto et al. “Design and validation of a foldable and photovoltaic wide-field epiretinal prosthesis”. en. In: *Nature Communications* 9.1 (Mar. 2018). Publisher: Nature Publishing Group, p. 992. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03386-7. URL: <https://www.nature.com/articles/s41467-018-03386-7> (visited on 12/02/2024).
- [125] N.M. Jansonius et al. “A mathematical description of nerve fiber bundle trajectories and their variability in the human retina”. en. In: *Vision Research* 49.17 (Aug. 2009), pp. 2157–2163. ISSN: 00426989. DOI: 10.1016/j.visres.2009.04.029. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0042698909002922> (visited on 02/03/2023).
- [126] Michael Beyeler et al. “Model-Based Recommendations for Optimal Surgical Placement of Epiretinal Implants”. In: (2019). DOI: 10.1007/978-3-030-32254-0_44. URL: https://doi.org/10.1007/978-3-030-32254-0_44.
- [127] Scott H. Greenwald et al. “Brightness as a Function of Current Amplitude in Human Retinal Electrical Stimulation”. en. In: *Investigative Ophthalmology & Visual Science* 50.11 (Nov. 2009), p. 5017. ISSN: 1552-5783. DOI: 10.1167/iovs.08-2897. URL: <http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.08-2897> (visited on 02/03/2023).
- [128] Dorsa Haji Ghaffari et al. “The effect of waveform asymmetry on perception with epiretinal prostheses”. en. In: *Journal of Neural Engineering* 17.4 (July 2020). Publisher: IOP Publishing, p. 045009. ISSN: 1741-2552. DOI: 10.1088/1741-2552/aba07e. URL: <https://dx.doi.org/10.1088/1741-2552/aba07e> (visited on 08/26/2025).

- [129] Chris Sekirnjak et al. “Electrical Stimulation of Mammalian Retinal Ganglion Cells With Multielectrode Arrays”. en. In: *Journal of Neurophysiology* 95.6 (June 2006), pp. 3311–3327. ISSN: 0022-3077, 1522-1598. DOI: 10.1152/jn.01168.2005. URL: <https://www.physiology.org/doi/10.1152/jn.01168.2005> (visited on 08/26/2025).
- [130] A.K. Ahuja et al. “An *In Vitro* Model of a Retinal Prosthesis”. en. In: *IEEE Transactions on Biomedical Engineering* 55.6 (June 2008), pp. 1744–1753. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2008.919126. URL: <http://ieeexplore.ieee.org/document/4457864/> (visited on 10/11/2020).
- [131] William H. Bosking et al. “Saturation in Phosphene Size with Increasing Current Levels Delivered to Human Visual Cortex”. en. In: *The Journal of Neuroscience* 37.30 (July 2017), pp. 7188–7197. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2896-16.2017. URL: <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2896-16.2017> (visited on 10/13/2021).
- [132] Caterina Cinel, Davide Valeriani, and Riccardo Poli. “Neurotechnologies for Human Cognitive Augmentation: Current State of the Art and Future Prospects”. In: *Frontiers in Human Neuroscience* 13 (Jan. 2019), p. 13. ISSN: 1662-5161. DOI: 10.3389/fnhum.2019.00013. URL: <https://www.frontiersin.org/article/10.3389/fnhum.2019.00013/full> (visited on 05/16/2023).
- [133] Blake S. Wilson et al. “Better speech recognition with cochlear implants”. en. In: *Nature* 352.6332 (July 1991), pp. 236–238. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/352236a0. URL: <https://www.nature.com/articles/352236a0> (visited on 05/17/2023).
- [134] Daniel W. Tan et al. “A neural interface provides long-term stable natural touch perception”. In: *Science Translational Medicine* 6.257 (Oct. 2014). Publisher:

- American Association for the Advancement of Science, 257ra138–257ra138. DOI: 10.1126/scitranslmed.3008669. URL: <https://www.science.org/doi/full/10.1126/scitranslmed.3008669> (visited on 03/16/2025).
- [135] Bingni W. Brunton and Michael Beyeler. “Data-driven models in human neuroscience and neuroengineering”. en. In: *Current Opinion in Neurobiology*. Computational Neuroscience 58 (Oct. 2019), pp. 21–29. ISSN: 0959-4388. DOI: 10.1016/j.conb.2019.06.008. URL: <https://www.sciencedirect.com/science/article/pii/S0959438818302502> (visited on 10/10/2022).
- [136] David Montes de Oca Zapiain, James A. Stewart, and Rémi Dingreville. “Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods”. en. In: *npj Computational Materials* 7.1 (Jan. 2021). Publisher: Nature Publishing Group, pp. 1–11. ISSN: 2057-3960. DOI: 10.1038/s41524-020-00471-8. URL: <https://www.nature.com/articles/s41524-020-00471-8> (visited on 03/16/2025).
- [137] Mohammad Amin Nabian and Hadi Meidani. “A Deep Neural Network Surrogate for High-Dimensional Random Partial Differential Equations”. In: *Probabilistic Engineering Mechanics* 57 (July 2019). arXiv:1806.02957 [cs], pp. 14–25. ISSN: 02668920. DOI: 10.1016/j.probengmech.2019.05.001. URL: <http://arxiv.org/abs/1806.02957> (visited on 03/16/2025).
- [138] Stefanos Nikolopoulos, Ioannis Kalogeris, and Vissarion Papadopoulos. “Non-intrusive surrogate modeling for parametrized time-dependent partial differential equations using convolutional autoencoders”. In: *Engineering Applications of Artificial Intelligence* 109 (Mar. 2022), p. 104652. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2021.104652. URL: <https://www.sciencedirect.com/science/article/pii/S0952197621004541> (visited on 03/16/2025).

- [139] Gregg A. Tabot et al. “Restoring the sense of touch with a prosthetic hand through a brain interface”. en. In: *Proceedings of the National Academy of Sciences* 110.45 (Nov. 2013), pp. 18279–18284. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1221113110. URL: <https://pnas.org/doi/full/10.1073/pnas.1221113110> (visited on 05/17/2023).
- [140] Marco Capogrosso et al. “A brain–spine interface alleviating gait deficits after spinal cord injury in primates”. en. In: *Nature* 539.7628 (Nov. 2016). Publisher: Nature Publishing Group, pp. 284–288. ISSN: 1476-4687. DOI: 10.1038/nature20118. URL: <https://www.nature.com/articles/nature20118> (visited on 03/16/2025).
- [141] Christopher A. R. Chapman, Noah Goshi, and Erkin Seker. “Multifunctional Neural Interfaces for Closed-Loop Control of Neural Activity”. en. In: *Advanced Functional Materials* 28.12 (2018), p. 1703523. ISSN: 1616-3028. DOI: 10.1002/adfm.201703523. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.201703523> (visited on 03/16/2025).
- [142] Fabien B. Wagner et al. “Targeted neurotechnology restores walking in humans with spinal cord injury”. en. In: *Nature* 563.7729 (Nov. 2018). Publisher: Nature Publishing Group, pp. 65–71. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0649-2. URL: <https://www.nature.com/articles/s41586-018-0649-2> (visited on 03/16/2025).
- [143] Katarina Stingl et al. “Interim Results of a Multicenter Trial with the New Electronic Subretinal Implant Alpha AMS in 15 Patients Blind from Inherited Retinal Degenerations”. English. In: *Frontiers in Neuroscience* 11 (2017). Publisher: Frontiers. ISSN: 1662-453X. DOI: 10.3389/fnins.2017.00445. URL: <https://www.frontiersin.org/articles/10.3389/fnins.2017.00445/full> (visited on 05/28/2020).

- [144] Lewis Karapanos et al. “Functional Vision in the Real-World Environment With a Second-Generation (44-Channel) Suprachoroidal Retinal Prosthesis”. en. In: *Translational Vision Science & Technology* 10.10 (Aug. 2021). Publisher: The Association for Research in Vision and Ophthalmology, pp. 7–7. ISSN: 2164-2591. DOI: 10.1167/tvst.10.10.7. URL: <https://tvst.arvojournals.org/article.aspx?articleid=2776583> (visited on 08/18/2021).
- [145] Joseph F. Rizzo III et al. “Perceptual Efficacy of Electrical Stimulation of Human Retina with a Microelectrode Array during Short-Term Surgical Trials”. In: *Investigative Ophthalmology & Visual Science* 44.12 (Dec. 2003), pp. 5362–5369. ISSN: 1552-5783. DOI: 10.1167/iovs.02-0817. URL: <https://doi.org/10.1167/iovs.02-0817> (visited on 10/05/2021).
- [146] Robert E Marc et al. “Neural remodeling in retinal degeneration”. In: *Progress in Retinal and Eye Research* 22.5 (Sept. 2003), pp. 607–655. ISSN: 1350-9462. DOI: 10.1016/S1350-9462(03)00039-9. URL: <https://www.sciencedirect.com/science/article/pii/S1350946203000399> (visited on 03/16/2025).
- [147] Zhou Wang and Alan C. Bovik. “Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures”. In: *IEEE Signal Processing Magazine* 26.1 (Jan. 2009). Conference Name: IEEE Signal Processing Magazine, pp. 98–117. ISSN: 1558-0792. DOI: 10.1109/MSP.2008.930649. URL: <https://ieeexplore.ieee.org/abstract/document/4775883> (visited on 03/16/2025).
- [148] Guangtao Zhai and Xiongkuo Min. “Perceptual image quality assessment: a survey”. en. In: *Science China Information Sciences* 63.11 (Nov. 2020), p. 211301. ISSN: 1674-733X, 1869-1919. DOI: 10.1007/s11432-019-2757-1. URL: <https://link.springer.com/10.1007/s11432-019-2757-1> (visited on 03/03/2022).

- [149] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. en. In: *Proceedings of the 32nd International Conference on Machine Learning*. ISSN: 1938-7228. PMLR, June 2015, pp. 448–456. URL: <https://proceedings.mlr.press/v37/ioffe15.html> (visited on 03/16/2025).
- [150] Michael F. Dorman et al. “Acoustic Simulations of Combined Electric and Acoustic Hearing (EAS)”. en-US. In: *Ear and Hearing* 26.4 (Aug. 2005), p. 371. ISSN: 1538-4667. URL: https://journals.lww.com/ear-hearing/fulltext/2005/08000/Acoustic_Simulations_of_Combined_Electric_and.1.aspx (visited on 03/16/2025).
- [151] Mario A. Svirsky et al. “Validation of acoustic models of auditory neural prostheses”. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. ISSN: 2379-190X. May 2013, pp. 8629–8633. DOI: 10.1109/ICASSP.2013.6639350. URL: <https://ieeexplore.ieee.org/abstract/document/6639350> (visited on 03/16/2025).
- [152] M.F. Dorman et al. “Simulations of combined acoustic/electric hearing”. In: *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*. Vol. 3. ISSN: 1094-687X. Sept. 2003, 1999–2001 Vol.3. DOI: 10.1109/IEMBS.2003.1280125. URL: <https://ieeexplore.ieee.org/abstract/document/1280125> (visited on 03/16/2025).
- [153] Michael F. Dorman, Philipos C. Loizou, and Dawne Rainey. “Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs”. In: *The Journal of the Acoustical Society of America*

- 102.4 (Oct. 1997), pp. 2403–2411. ISSN: 0001-4966. DOI: 10.1121/1.419603. URL: <https://doi.org/10.1121/1.419603> (visited on 03/16/2025).
- [154] William B. Cooper, Emily Tobey, and Philipos C. Loizou. “Music Perception by Cochlear Implant and Normal Hearing Listeners as Measured by the Montreal Battery for Evaluation of Amusia”. en-US. In: *Ear and Hearing* 29.4 (Aug. 2008), p. 618. ISSN: 1538-4667. DOI: 10.1097/AUD.0b013e318174e787. URL: https://journals.lww.com/ear-hearing/fulltext/2008/08000/Perception_of_Rhythmic_and_Sequential_Pitch.12.aspx (visited on 03/16/2025).
- [155] Philipos C. Loizou et al. “Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution”. In: *The Journal of the Acoustical Society of America* 108.5 (Nov. 2000), pp. 2377–2387. ISSN: 0001-4966. DOI: 10.1121/1.1317557. URL: <https://doi.org/10.1121/1.1317557> (visited on 03/16/2025).
- [156] Hannes P. Saal et al. “Simulating tactile signals from the whole hand with millisecond precision”. en. In: *Proceedings of the National Academy of Sciences* 114.28 (July 2017). ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1704856114. URL: <https://pnas.org/doi/full/10.1073/pnas.1704856114> (visited on 05/23/2023).
- [157] Elizaveta V. Okorokova, Qinqu He, and Sliman J. Bensmaia. “Biomimetic encoding model for restoring touch in bionic hands through a nerve interface”. en. In: *Journal of Neural Engineering* 15.6 (Oct. 2018). Publisher: IOP Publishing, p. 066033. ISSN: 1741-2552. DOI: 10.1088/1741-2552/aae398. URL: <https://dx.doi.org/10.1088/1741-2552/aae398> (visited on 05/17/2023).
- [158] Douglas J. Weber, Rebecca Friesen, and Lee E. Miller. “Interfacing the Somatosensory System to Restore Touch and Proprioception: Essential Considerations”. en. In: *Journal of Motor Behavior* 44.6 (Nov. 2012), pp. 403–418. ISSN: 0022-2895, 1940-

BIBLIOGRAPHY

1027. DOI: 10.1080/00222895.2012.735283. URL: <http://www.tandfonline.com/doi/abs/10.1080/00222895.2012.735283> (visited on 03/16/2025).
- [159] Sung Soo Kim. “Predicting the Timing of Spikes Evoked by Tactile Stimulation of the Hand”. In: *J. Neurophysiology* (2010). URL: <https://journals.physiology.org/doi/epdf/10.1152/jn.00187.2010> (visited on 03/16/2025).
- [160] Milana P. Mileusnic et al. “Mathematical Models of Proprioceptors. I. Control and Transduction in the Muscle Spindle”. In: *Journal of Neurophysiology* 96.4 (Oct. 2006). Publisher: American Physiological Society, pp. 1772–1788. ISSN: 0022-3077. DOI: 10.1152/jn.00868.2005. URL: <https://journals.physiology.org/doi/full/10.1152/jn.00868.2005> (visited on 05/23/2023).
- [161] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. arXiv:1603.04467 [cs]. Mar. 2016. DOI: 10.48550/arXiv.1603.04467. URL: <http://arxiv.org/abs/1603.04467> (visited on 03/16/2025).
- [162] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv:1409.1556 [cs]. Apr. 2015. DOI: 10.48550/arXiv.1409.1556. URL: <http://arxiv.org/abs/1409.1556> (visited on 03/16/2025).
- [163] Yijun Li et al. “Universal Style Transfer via Feature Transforms”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/49182f81e6a13cf5eaa496d51fea6406-Paper.pdf.
- [164] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE, June 2009, pp. 248–255. ISBN: 978-1-4244-3992-8. DOI: 10.1109/CVPR.2009.

5206848. URL: <https://ieeexplore.ieee.org/document/5206848/> (visited on 05/17/2023).
- [165] Sylvain Paris, Samuel W Hasinoff, and Jan Kautz. “Local Laplacian Filters: Edge-aware Image Processing with a Laplacian Pyramid”. en. In: (2011).
- [166] Tsung-Yi Lin et al. *Microsoft COCO: Common Objects in Context*. arXiv:1405.0312 [cs]. Feb. 2015. DOI: 10.48550/arXiv.1405.0312. URL: <http://arxiv.org/abs/1405.0312> (visited on 03/16/2025).
- [167] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. ISSN: 1533-7928. URL: <http://jmlr.org/papers/v9/vandermaaten08a.html> (visited on 08/27/2025).
- [168] Eli Peli. “Testing Vision Is Not Testing For Vision”. In: *Translational Vision Science & Technology* 9.13 (Dec. 2020), p. 32. ISSN: 2164-2591. DOI: 10.1167/tvst.9.13.32. URL: <https://doi.org/10.1167/tvst.9.13.32> (visited on 08/08/2023).
- [169] Florian Tramèr et al. *Ensemble Adversarial Training: Attacks and Defenses*. arXiv:1705.07204 [stat]. Apr. 2020. DOI: 10.48550/arXiv.1705.07204. URL: <http://arxiv.org/abs/1705.07204> (visited on 03/16/2025).
- [170] Alim Louis Benabid. “Deep brain stimulation for Parkinson’s disease”. In: *Current Opinion in Neurobiology* 13.6 (Dec. 2003), pp. 696–706. ISSN: 0959-4388. DOI: 10.1016/j.conb.2003.11.001. URL: <https://www.sciencedirect.com/science/article/pii/S0959438803001739> (visited on 03/16/2025).
- [171] Bardia Bozorgzadeh et al. “Neurochemostat: A Neural Interface SoC With Integrated Chemometrics for Closed-Loop Regulation of Brain Dopamine”. en. In: *IEEE Transactions on Biomedical Circuits and Systems* 10.3 (June 2016), pp. 654–

667. ISSN: 1932-4545, 1940-9990. DOI: 10.1109/TBCAS.2015.2453791. URL: <http://ieeexplore.ieee.org/document/7270344/> (visited on 03/16/2025).
- [172] Fabrizio Grani et al. "Toward a personalized closed-loop stimulation of the visual cortex: Advances and challenges". In: *Frontiers in Cellular Neuroscience* 16 (Dec. 2022), p. 1034270. ISSN: 1662-5102. DOI: 10.3389/fncel.2022.1034270. URL: <https://www.frontiersin.org/articles/10.3389/fncel.2022.1034270/full> (visited on 09/26/2024).
- [173] Sumedh Sopan Nagrale et al. "In silico development and validation of Bayesian methods for optimizing deep brain stimulation to enhance cognitive control". en. In: *Journal of Neural Engineering* (Apr. 2023). ISSN: 1741-2560, 1741-2552. DOI: 10.1088/1741-2552/acd0d5. URL: <https://iopscience.iop.org/article/10.1088/1741-2552/acd0d5> (visited on 05/14/2023).
- [174] Ghazaleh Soleimani et al. *Closing the loop between brain and electrical stimulation: Towards precision neuromodulation treatments*. en. preprint. PsyArXiv, July 2022. DOI: 10.31234/osf.io/mxy85. URL: <https://osf.io/mxy85> (visited on 05/14/2023).
- [175] Bobak Shahriari et al. "Taking the Human Out of the Loop: A Review of Bayesian Optimization". en. In: *Proceedings of the IEEE* 104.1 (Jan. 2016), pp. 148–175. ISSN: 0018-9219, 1558-2256. DOI: 10.1109/JPROC.2015.2494218. URL: <http://ieeexplore.ieee.org/document/7352306/> (visited on 02/03/2023).
- [176] Dorsa Haji Ghaffari et al. "Real-Time Optimization of Retinal Ganglion Cell Spatial Activity in Response to Epiretinal Stimulation". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 2733–2741. ISSN: 1534-4320, 1558-0210. DOI: 10.1109/TNSRE.2021.3138297. URL: <http://ieeexplore.ieee.org/document/9662355/> (visited on 05/17/2023).

- [177] Sasidhar S. Madugula et al. *Focal Electrical Stimulation of Human Retinal Ganglion Cells for Vision Restoration*. en. preprint. Neuroscience, Aug. 2020. DOI: 10.1101/2020.08.23.263608. URL: <http://biorxiv.org/lookup/doi/10.1101/2020.08.23.263608> (visited on 05/17/2023).
- [178] Nishal P. Shah et al. *Precise control of neural activity using dynamically optimized electrical stimulation*. en. July 2022. DOI: 10.1101/2022.07.26.501643. URL: <http://biorxiv.org/lookup/doi/10.1101/2022.07.26.501643> (visited on 09/26/2024).
- [179] Maureen van der Grinten et al. *Biologically plausible phosphene simulation for the differentiable optimization of visual cortical prostheses*. en. Pages: 2022.12.23.521749 Section: New Results. Dec. 2022. DOI: 10.1101/2022.12.23.521749. URL: <https://www.biorxiv.org/content/10.1101/2022.12.23.521749v1> (visited on 05/17/2023).
- [180] Eric Brochu, Tyson Brochu, and Nando De Freitas. “A Bayesian Interactive Optimization Approach to Procedural Animation Design”. en. In: *Eurographics/ACM SIGGRAPH Symposium on Computer Animation* (2010). Artwork Size: 10 pages ISBN: 9783905674279 Publisher: The Eurographics Association, 10 pages. ISSN: 1727-5288. DOI: 10.2312/SCA/SCA10/103-112. URL: <http://diglib.eg.org/handle/10.2312/SCA.SCA10.103-112> (visited on 05/17/2023).
- [181] Javier Gonzalez et al. *Preferential Bayesian Optimization*. en. arXiv:1704.03651 [stat]. Apr. 2017. URL: <http://arxiv.org/abs/1704.03651> (visited on 02/03/2023).
- [182] Thomas P. Minka. “Expectation Propagation for approximate Bayesian inference”. In: (2013). Publisher: arXiv Version Number: 1. DOI: 10.48550/ARXIV.1301.2294. URL: <https://arxiv.org/abs/1301.2294> (visited on 05/17/2023).

- [183] Matthias Seeger. “Expectation Propagation for Exponential Families”. en. In: (2008).
- [184] Roman Garnett. *Bayesian optimization*. Cambridge, United Kingdom ; New York, NY: Cambridge University Press, 2023. ISBN: 978-1-108-42578-0.
- [185] Romy Lorenz et al. “Efficiently searching through large tACS parameter spaces using closed-loop Bayesian optimization”. en. In: *Brain Stimulation* 12.6 (Nov. 2019), pp. 1484–1489. ISSN: 1935861X. DOI: 10.1016/j.brs.2019.07.003. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1935861X19302852> (visited on 02/03/2023).
- [186] Zixi Zhao et al. “Optimization of Spinal Cord Stimulation Using Bayesian Preference Learning and Its Validation”. en. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 1987–1997. ISSN: 1534-4320, 1558-0210. DOI: 10.1109/TNSRE.2021.3113636. URL: <https://ieeexplore.ieee.org/document/9540877/> (visited on 02/03/2023).
- [187] Tristan Fauvel and Matthew Chalk. “Human-in-the-loop optimization of visual prosthetic stimulation”. en. In: *Journal of Neural Engineering* 19.3 (June 2022). Publisher: IOP Publishing, p. 036038. ISSN: 1741-2552. DOI: 10.1088/1741-2552/ac7615. URL: <https://dx.doi.org/10.1088/1741-2552/ac7615> (visited on 05/10/2023).
- [188] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: (2014). Publisher: arXiv Version Number: 9. DOI: 10.48550/ARXIV.1412.6980. URL: <https://arxiv.org/abs/1412.6980> (visited on 05/17/2023).
- [189] Neil Houlsby et al. *Bayesian Active Learning for Classification and Preference Learning*. en. arXiv:1112.5745 [cs, stat]. Dec. 2011. URL: <http://arxiv.org/abs/1112.5745> (visited on 02/03/2023).

- [190] Wei Chu and Zoubin Ghahramani. “Preference learning with Gaussian processes”. en. In: *Proceedings of the 22nd international conference on Machine learning - ICML '05*. Bonn, Germany: ACM Press, 2005, pp. 137–144. ISBN: 978-1-59593-180-1. DOI: 10.1145/1102351.1102369. URL: <http://portal.acm.org/citation.cfm?doid=1102351.1102369> (visited on 02/03/2023).
- [191] Tristan Fauvel and Matthew Chalk. *Efficient Exploration in Binary and Preferential Bayesian Optimization*. arXiv:2110.09361 [cs, math, stat]. Oct. 2021. DOI: 10.48550/arXiv.2110.09361. URL: <http://arxiv.org/abs/2110.09361> (visited on 05/14/2023).
- [192] Glen Brier. “Verification of forecasts expressed in terms of probability”. en. In: *Monthly Weather Review*. Google-Books-ID: jnbpAAAAMAAJ. War Department, Office of the Chief Signal Officer, 1950.
- [193] Nick Barnes et al. “Vision function testing for a suprachoroidal retinal prosthesis: effects of image filtering”. en. In: *Journal of Neural Engineering* 13.3 (June 2016), p. 036013. ISSN: 1741-2560, 1741-2552. DOI: 10.1088/1741-2560/13/3/036013. URL: <https://iopscience.iop.org/article/10.1088/1741-2560/13/3/036013> (visited on 02/03/2023).
- [194] Lan Yue et al. “Ten-Year Follow-up of a Blind Patient Chronically Implanted with Epiretinal Prosthesis Argus I”. In: *Ophthalmology* 122.12 (Dec. 2015), 2545–2552.e1. ISSN: 0161-6420. DOI: 10.1016/j.opthta.2015.08.008. URL: <http://www.sciencedirect.com/science/article/pii/S016164201500785X> (visited on 10/15/2019).
- [195] Fotios Drakopoulos and Sarah Verhulst. “A Neural-Network Framework for the Design of Individualised Hearing-Loss Compensation”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023). Conference Name: IEEE/ACM

- Transactions on Audio, Speech, and Language Processing, pp. 1–15. ISSN: 2329-9304. DOI: 10.1109/TASLP.2023.3282093.
- [196] Ekta Prashnani et al. “PieAPP: Perceptual Image-Error Assessment Through Pairwise Preference”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 1808–1817. URL: https://openaccess.thecvf.com/content_cvpr_2018/html/Prashnani_PieAPP_Perceptual_Image-Error_CVPR_2018_paper.html (visited on 10/10/2022).
- [197] Lily M. Turkstra et al. “Assistive technology use in domestic activities by people who are blind”. en. In: *Scientific Reports* 15.1 (Mar. 2025). Publisher: Nature Publishing Group, p. 7486. ISSN: 2045-2322. DOI: 10.1038/s41598-025-91755-w. URL: <https://www.nature.com/articles/s41598-025-91755-w> (visited on 09/01/2025).
- [198] Jacob Granley et al. “Beyond Sight: Probing Alignment Between Image Models and Blind V1”. In: arXiv:2403.12990 [q-bio]. arXiv, Mar. 2024. DOI: 10.48550/arXiv.2403.12990. URL: <http://arxiv.org/abs/2403.12990> (visited on 09/01/2025).
- [199] Nicholas A. Lesica et al. “Harnessing the power of artificial intelligence to transform hearing healthcare and research”. en. In: *Nature Machine Intelligence* 3.10 (Oct. 2021). Publisher: Nature Publishing Group, pp. 840–849. ISSN: 2522-5839. DOI: 10.1038/s42256-021-00394-z. URL: <https://www.nature.com/articles/s42256-021-00394-z> (visited on 09/01/2025).
- [200] Matthew J Bryan, Linxing Preston Jiang, and Rajesh P N Rao. “Neural co-processors for restoring brain function: results from a cortical model of grasping”. en. In: *Journal of Neural Engineering* 20.3 (May 2023). Publisher: IOP Publishing, p. 036004. ISSN: 1741-2552. DOI: 10.1088/1741-2552/accaa9. URL: <https://dx.doi.org/10.1088/1741-2552/accaa9> (visited on 09/01/2025).

- [201] Mark H. Histed, Vincent Bonin, and R. Clay Reid. “Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation”. eng. In: *Neuron* 63.4 (Aug. 2009), pp. 508–522. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2009.07.016.
- [202] Silvestro Micera et al. *A computational model to design wide field-of-view optic nerve neuroprostheses*. Aug. 2023. DOI: 10.21203/rs.3.rs-3218482/v1. URL: <https://www.researchsquare.com/article/rs-3218482/v1> (visited on 09/26/2024).
- [203] Praful K. Vasireddy et al. “Efficient Modeling and Calibration of Multi-Electrode Stimuli for Epiretinal Implants”. In: *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*. ISSN: 1948-3554. Apr. 2023, pp. 1–4. DOI: 10.1109/NER52421.2023.10123907. URL: <https://ieeexplore.ieee.org/document/10123907> (visited on 09/26/2024).
- [204] Sina Tafazoli et al. “Learning to control the brain through adaptive closed-loop patterned stimulation”. en. In: *Journal of Neural Engineering* 17.5 (Oct. 2020), p. 056007. ISSN: 1741-2552. DOI: 10.1088/1741-2552/abb860. URL: <https://iopscience.iop.org/article/10.1088/1741-2552/abb860> (visited on 10/10/2022).
- [205] Xing Chen et al. “Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex”. en. In: *Science* 370.6521 (Dec. 2020), pp. 1191–1196. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abd7435. URL: <https://www.science.org/doi/10.1126/science.abd7435> (visited on 09/26/2024).
- [206] Paolo Papale, Daniela De Luca, and Pieter R. Roelfsema. *Deep generative networks reveal the tuning of neurons in IT and predict their influence on visual perception*. en. Pages: 2024.10.09.617382 Section: New Results. Oct. 2024. DOI: 10.1101/2024.

BIBLIOGRAPHY

- 10.09.617382. URL: <https://www.biorxiv.org/content/10.1101/2024.10.09.617382v1> (visited on 08/26/2025).
- [207] Fabrizio Grani et al. “ADVANTAGES OF BIDIRECTIONAL CORTICAL VISUAL PROSTHESES”. en. In: *IBRO Neuroscience Reports*. Vol. 15. Oct. 2023, S948–S949. DOI: 10.1016/j.ibneur.2023.08.2003. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2667242123020651> (visited on 09/26/2024).
- [208] Tim Allison-Walker et al. “Local field potential phase modulates the evoked response to electrical stimulation in visual cortex”. en. In: *Journal of Neural Engineering* 22.1 (Jan. 2025). Publisher: IOP Publishing, p. 016009. ISSN: 1741-2552. DOI: 10.1088/1741-2552/ada828. URL: <https://dx.doi.org/10.1088/1741-2552/ada828> (visited on 08/26/2025).
- [209] Timothy John Allison-Walker et al. “Local field potential phase modulates neural responses to intracortical electrical stimulation”. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. ISSN: 2694-0604. July 2020, pp. 3521–3524. DOI: 10.1109/EMBC44109.2020.9176186. URL: <https://ieeexplore.ieee.org/abstract/document/9176186> (visited on 08/26/2025).
- [210] Laura Dugué, Philippe Marque, and Rufin VanRullen. “The Phase of Ongoing Oscillations Mediates the Causal Relation between Brain Excitation and Visual Perception”. en. In: *Journal of Neuroscience* 31.33 (Aug. 2011). Publisher: Society for Neuroscience Section: Brief Communications, pp. 11889–11893. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1161-11.2011. URL: <https://www.jneurosci.org/content/31/33/11889> (visited on 08/26/2025).

BIBLIOGRAPHY

- [211] Junior Leo, Ernest Ge, and Stotle Li. *Wasserstein Distance in Deep Learning*. en. SSRN Scholarly Paper. Rochester, NY, Feb. 2023. DOI: 10.2139/ssrn.4368733. URL: <https://papers.ssrn.com/abstract=4368733> (visited on 01/10/2025).
- [212] Pouya Bashivan, Kohitij Kar, and James J. DiCarlo. “Neural population control via deep image synthesis”. en. In: *Science* 364.6439 (May 2019), eaav9436. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aav9436. URL: <https://www.science.org/doi/10.1126/science.aav9436> (visited on 09/26/2024).
- [213] Edgar Y. Walker et al. “Inception loops discover what excites neurons most using deep predictive models”. en. In: *Nature Neuroscience* 22.12 (Dec. 2019). Publisher: Nature Publishing Group, pp. 2060–2065. ISSN: 1546-1726. DOI: 10.1038/s41593-019-0517-x. URL: <https://www.nature.com/articles/s41593-019-0517-x> (visited on 08/26/2025).
- [214] Aitor Morales-Gregorio et al. “Neural manifolds in V1 change with top-down signals from V4 targeting the foveal region”. English. In: *Cell Reports* 43.7 (July 2024). Publisher: Elsevier. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2024.114371. URL: [https://www.cell.com/cell-reports/abstract/S2211-1247\(24\)00699-5](https://www.cell.com/cell-reports/abstract/S2211-1247(24)00699-5) (visited on 08/26/2025).
- [215] Christopher Langdon, Mikhail Genkin, and Tatiana A. Engel. “A unifying perspective on neural manifolds and circuits for cognition”. en. In: *Nature Reviews Neuroscience* 24.6 (June 2023). Publisher: Nature Publishing Group, pp. 363–377. ISSN: 1471-0048. DOI: 10.1038/s41583-023-00693-x. URL: <https://www.nature.com/articles/s41583-023-00693-x> (visited on 08/26/2025).
- [216] David J. Bartholomew, Martin Knott, and Irini Moustaki. *Latent Variable Models and Factor Analysis: A Unified Approach*. en. Google-Books-ID: 5YxvE1S56DwC. John Wiley & Sons, June 2011. ISBN: 978-1-119-97370-6.

BIBLIOGRAPHY

- [217] Patrick T. Sadtler et al. “Neural constraints on learning”. en. In: *Nature* 512.7515 (Aug. 2014). Publisher: Nature Publishing Group, pp. 423–426. ISSN: 1476-4687. DOI: 10.1038/nature13665. URL: <https://www.nature.com/articles/nature13665> (visited on 08/26/2025).
- [218] Carsen Stringer et al. “High-dimensional geometry of population responses in visual cortex”. en. In: *Nature* 571.7765 (July 2019). Publisher: Nature Publishing Group, pp. 361–365. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1346-5. URL: <https://www.nature.com/articles/s41586-019-1346-5> (visited on 08/26/2025).
- [219] Richard A Normann and Eduardo Fernandez. “Clinical applications of penetrating neural interfaces and Utah Electrode Array technologies”. In: *Journal of Neural Engineering* 13.6 (Dec. 2016), p. 061003. ISSN: 1741-2560, 1741-2552. DOI: 10.1088/1741-2560/13/6/061003. URL: <https://iopscience.iop.org/article/10.1088/1741-2560/13/6/061003> (visited on 09/26/2024).
- [220] Alessandra Rocca et al. “Robot-assisted implantation of a microelectrode array in the occipital lobe as a visual prosthesis: technical note”. In: *Journal of Neurosurgery* 140.4 (Apr. 2024), pp. 1169–1176. ISSN: 0022-3085, 1933-0693. DOI: 10.3171/2023.8.JNS23772. URL: <https://thejns.org/view/journals/j-neurosurg/140/4/article-p1169.xml> (visited on 09/26/2024).
- [221] Eric M. Trautmann et al. “Accurate Estimation of Neural Population Dynamics without Spike Sorting”. English. In: *Neuron* 103.2 (July 2019). Publisher: Elsevier, 292–308.e4. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2019.05.003. URL: [https://www.cell.com/neuron/abstract/S0896-6273\(19\)30428-3](https://www.cell.com/neuron/abstract/S0896-6273(19)30428-3) (visited on 08/26/2025).
- [222] Nur Ahmadi, Timothy G Constandinou, and Christos-Savvas Bouganis. “Robust and accurate decoding of hand kinematics from entire spiking activity using deep

- learning”. en. In: *Journal of Neural Engineering* 18.2 (Feb. 2021). Publisher: IOP Publishing, p. 026011. ISSN: 1741-2552. DOI: 10.1088/1741-2552/abde8a. URL: <https://dx.doi.org/10.1088/1741-2552/abde8a> (visited on 02/11/2025).
- [223] Andrew Aoun et al. “Beyond correlation: optimal transport metrics for characterizing representational stability and remapping in neurons encoding spatial memory”. English. In: *Frontiers in Cellular Neuroscience* 17 (Jan. 2024). Publisher: Frontiers. ISSN: 1662-5102. DOI: 10.3389/fncel.2023.1273283. URL: <https://www.frontiersin.org/journals/cellular-neuroscience/articles/10.3389/fncel.2023.1273283/full> (visited on 08/26/2025).
- [224] Bernard Rosner, Robert J. Glynn, and Mei-Ling T. Lee. “The Wilcoxon signed rank test for paired comparisons of clustered data”. eng. In: *Biometrics* 62.1 (Mar. 2006), pp. 185–192. ISSN: 0006-341X. DOI: 10.1111/j.1541-0420.2005.00389.x.
- [225] Yoav Benjamini and Daniel Yekutieli. “The control of the false discovery rate in multiple testing under dependency”. In: *The Annals of Statistics* 29.4 (Aug. 2001). Publisher: Institute of Mathematical Statistics, pp. 1165–1188. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1013699998. URL: <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-4/The-control-of-the-false-discovery-rate-in-multiple-testing/10.1214/aos/1013699998.full> (visited on 01/10/2025).
- [226] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. en. In: *MACHINE LEARNING IN PYTHON* (2011).
- [227] Ilya Loshchilov and Frank Hutter. “Decoupled Weight Decay Regularization”. In: *International Conference on Learning Representations*. arXiv:1711.05101 [cs, math]. arXiv, Jan. 2019. URL: <http://arxiv.org/abs/1711.05101> (visited on 10/10/2022).

- [228] Paola Vagni et al. “POLYRETINA restores light responses in vivo in blind Göttingen minipigs”. en. In: *Nature Communications* 13.1 (June 2022). Number: 1 Publisher: Nature Publishing Group, p. 3678. ISSN: 2041-1723. DOI: 10.1038/s41467-022-31180-z. URL: <https://www.nature.com/articles/s41467-022-31180-z> (visited on 05/23/2023).
- [229] Yuchen Hou et al. “Axonal stimulation affects the linear summation of single-point perception in three Argus II users”. en. In: *Journal of Neural Engineering* 21.2 (Apr. 2024). Publisher: IOP Publishing, p. 026031. ISSN: 1741-2552. DOI: 10.1088/1741-2552/ad31c4. URL: <https://dx.doi.org/10.1088/1741-2552/ad31c4> (visited on 09/06/2025).
- [230] Matthias Seeger. “Gaussian processes for machine learning”. In: *International Journal of Neural Systems* 14.02 (Apr. 2004). Publisher: World Scientific Publishing Co., pp. 69–106. ISSN: 0129-0657. DOI: 10.1142/S0129065704001899. URL: <https://www.worldscientific.com/doi/abs/10.1142/S0129065704001899> (visited on 05/23/2023).
- [231] Kenneth H. Louie et al. “Semi-automated approaches to optimize deep brain stimulation parameters in Parkinson’s disease”. en. In: *Journal of NeuroEngineering and Rehabilitation* 18.1 (Dec. 2021), p. 83. ISSN: 1743-0003. DOI: 10.1186/s12984-021-00873-9. URL: <https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-021-00873-9> (visited on 02/03/2023).
- [232] Maegan Tucker et al. *Preference-Based Learning for User-Guided HZD Gait Generation on Bipedal Walking Robots*. en. arXiv:2011.05424 [cs]. Mar. 2021. URL: <http://arxiv.org/abs/2011.05424> (visited on 02/03/2023).
- [233] Jens Brehm Bagger Nielsen, Jakob Nielsen, and Jan Larsen. “Perception-Based Personalization of Hearing Aids Using Gaussian Processes and Active Learning”.

- In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23.1 (Jan. 2015). Conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing, pp. 162–173. ISSN: 2329-9304. DOI: 10.1109/TASLP.2014.2377581.
- [234] Alessio Benavoli, Dario Azzimonti, and Dario Piga. “Preferential Bayesian optimisation with skew gaussian processes”. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. GECCO ’21. New York, NY, USA: Association for Computing Machinery, July 2021, pp. 1842–1850. ISBN: 978-1-4503-8351-6. DOI: 10.1145/3449726.3463128. URL: <https://dl.acm.org/doi/10.1145/3449726.3463128> (visited on 05/22/2023).
- [235] Galen Pogoncheff et al. *Explainable Machine Learning Predictions of Perceptual Sensitivity for Retinal Prostheses*. en. Pages: 2023.02.09.23285633. Feb. 2023. DOI: 10.1101/2023.02.09.23285633. URL: <https://www.medrxiv.org/content/10.1101/2023.02.09.23285633v1> (visited on 05/23/2023).